

Complex martingales and asymptotic enumeration

Mikhail Isaev^{*†} and Brendan D. McKay^{*}

^{*}Research School of Computer Science
Australian National University
Canberra ACT 2601, Australia

[†]Moscow Institute of Physics and Technology
Dolgoprudny, 141700, Russia

isaev.m.i@gmail.com, brendan.mckay@anu.edu.au

Abstract

Many enumeration problems in combinatorics, including such fundamental questions as the number of regular graphs, can be expressed as high-dimensional complex integrals. Motivated by the need for a systematic study of the asymptotic behaviour of such integrals, we establish explicit bounds on the exponentials of complex martingales. Those bounds applied to the case of truncated normal distributions are precise enough to include and extend many enumerative results of Barvinok, Canfield, Gao, Greenhill, Hartigan, Isaev, McKay, Wang, Wormald, and others. Our method applies to sums as well as integrals.

As a first illustration of the power of our theory, we considerably strengthen existing results on the relationship between random graphs or bipartite graphs with specified degrees and the so-called β -model of random graphs with independent edges, which is equivalent to the Rasch model in the bipartite case.

^{*}Research supported by the Australian Research Council.

1 Introduction

A large number of combinatorial enumeration problems can be expressed in terms of high-dimensional integrals, often, but not always, resulting from Fourier inversion applied to a multivariable generating function.

To illustrate what we mean, here are two examples. The number of undirected simple graphs with degrees d_1, \dots, d_n is given by

$$\frac{1}{(2\pi i)^n} \oint \cdots \oint \frac{\prod_{1 \leq j < k \leq n} (1 + z_j z_k)}{z_1^{d_1+1} \cdots z_n^{d_n+1}} dz_1 \cdots dz_n, \quad (1.1)$$

while the number of $m \times n$ nonnegative integer matrices (contingency tables) with row sums r_1, \dots, r_m and column sums c_1, \dots, c_n is given by

$$\frac{1}{(2\pi i)^{m+n}} \oint \cdots \oint \frac{\prod_{1 \leq j \leq m, 1 \leq k \leq n} (1 - w_j z_k)^{-1}}{w_1^{r_1+1} \cdots w_m^{r_m+1} z_1^{c_1+1} \cdots z_n^{c_n+1}} dw_1 \cdots dw_m dz_1 \cdots dz_n, \quad (1.2)$$

where each contour encloses the origin once anticlockwise. Although explicit evaluation of such integrals is rarely possible, under some circumstances asymptotic estimation is tractable. This was first achieved by McKay and Wormald in 1990, for (1.1) in the case of degree sequences not far from regular [36] and some classes of digraphs that include regular tournaments [32].

Since then, many other examples have appeared that include classes of 0-1 matrices [3, 5, 7, 8, 15, 31, 38]; directed graphs by degree sequence [14, 15, 32, 35, 41, 42]; eulerian digraphs [21, 43]; eulerian circuits [19, 20, 22, 34]; types of integer matrices [4, 9, 33]; and multiple other problems [6, 27, 37].

Estimation of integrals like (1.1) and (1.2) involves several steps, none of them trivial.

- (a) Choose as contours circles $r_j e^{i\theta_j}$ whose radii are chosen so that they pass together through the saddle-point (or close enough to it). This involves solving nonlinear equations or maximizing an entropy function.
- (b) Identify one or more small regions (in $\{\theta_j\}$ -space) in which the value of the integral is concentrated. This might be a small box enclosing the origin (as in (1.1)) or the neighbourhood of a low-dimensional subspace (as in (1.2)).
- (c) Within those small regions, approximate the integrand by a more tractable function and estimate its integral.

The present paper is motivated by step (c). The integrals that occur are typically of the

form

$$I = \int_B \exp(-\mathbf{x}^T A \mathbf{x} + f(\mathbf{x})) d\mathbf{x},$$

where B is a region containing the origin, A is a positive-semidefinite real matrix, and $f(\mathbf{x})$ is a function well-approximated by a truncated Taylor series with complex coefficients. The matrix A might not be of full rank.

Now let \mathbf{X} be a random variable whose distribution is given by the gaussian density $C \exp(-\mathbf{x}^T A \mathbf{x})$ truncated to domain B , where C is the normalising constant. Then we have

$$I = C^{-1} \mathbb{E} e^{f(\mathbf{X})},$$

so the problem is reduced to estimating $\mathbb{E} e^{f(\mathbf{X})}$. Our main aim is to make estimation of such integrals more systematic by providing some general theory about $\mathbb{E} e^{f(\mathbf{X})}$.

We will give bounds on $\mathbb{E} e^{f(\mathbf{X})}$ that are general and precise enough to cover and generalize all of the examples listed above and many more similar examples. In fact, we will not restrict ourselves to truncated gaussian measures or to functions f that are approximated by polynomials. Furthermore, both our measure and our functions f can be either smooth or discrete, allowing for sums as well as integrals.

1.1 Summary of the paper

Section 2 gives our main theorem in its most general form, providing explicit bounds on $\mathbb{E} e^{Z_n}$ when Z_0, \dots, Z_n is a complex martingale, based on properties of the martingale differences. Section 3 applies the martingale theorems to functions of independent random variables, via the Doob martingale. We also show how to bound the necessary parameters for smooth functions and how to handle vector measures whose components are independent only when the measure is rotated.

Section 4 considers the case of gaussian measures which are truncated to a finite region (usually a cuboid, perhaps intersected with a linear subspace). These are the theorems which can be applied directly to the enumeration problems we have surveyed. The cases of full-rank and non-full-rank gaussians are somewhat different. Finally in that section we give some lemmas useful for managing the quadratic forms which occur.

In Section 5 we demonstrate the power of our theorems using the example of graphs or bipartite graphs with given degrees. In each case, we allow degree sequences as general as those allowed by Barvinok and Hartigan [5], but we also allow a moderate number of forced and forbidden edges. This permits us to prove, in Section 5.3, that the corresponding β -

models are closer than previously known to the uniform model of random graphs with given degrees.

The Appendix collects some technical lemmas we need in the proofs.

1	Introduction	2
1.1	Summary of the paper	3
2	The exponential of a complex martingale	4
2.1	The diameter of a complex random variable	5
2.2	First order approach	10
2.3	Second order approach	12
3	Functions of independent random variables	15
3.1	Estimating the exponential	17
3.2	Smooth and transformed functions	20
4	Truncated gaussian measures	22
4.1	Truncated gaussian measures of full rank	25
4.2	Truncated gaussian measures of less than full rank	28
4.3	Example: regular tournaments	30
4.4	The case of weakly dependent components	32
5	Graphs with given degrees	35
5.1	General graphs	37
5.2	Bipartite graphs	41
5.3	Concentration near the β -model	46
6	Appendix	49

2 The exponential of a complex martingale

In this section we state and prove our theorems in their most general forms.

Let $\mathbf{P} = (\Omega, \mathcal{F}, P)$ be a probability space. We are interested in estimates for $\mathbb{E}e^Z$, where Z is a complex-valued random variable on \mathbf{P} . Such estimates for the case of real Z are commonplace as intermediate steps towards concentration inequalities, such as in the classical works of Hoeffding and McDiarmid [17, 30]. However, we seek $\mathbb{E}e^Z$ itself and few such intermediate results carry over unchanged to the complex case, perhaps fundamentally due to the non-convexity of the exponential function in the complex plane.

As our primary measure of spread of a complex random variable we use the diameter

of its essential support. This choice was inspired by its effective use (in the real case) by McDiarmid in analysing concentration of functions of independent random variables. A bound on $|f(\mathbf{x}') - f(\mathbf{x})|$ whenever \mathbf{x}, \mathbf{x}' differ only in the k -th position is, roughly speaking, the same as a bound on the diameter of the random variable $f(x_1, \dots, x_{k-1}, X_k, x_{k+1}, \dots, x_n)$ for constant $x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n$.

Note that having diameter α is weaker than being confined to a disk of diameter α . So while we could alternatively have generalised real intervals into complex disks, doing so would weaken our theorems.

In the next subsection we define the diameter formally, including a conditional version, and prove some properties that we will need. Then, in two further subsections, we use the diameter to bound the exponential of a complex martingale.

Recall that for complex random variables Z there are two types of squared variation commonly defined. The variance is

$$\text{Var } Z = \mathbb{E} |Z - \mathbb{E} Z|^2 = \mathbb{E} |Z|^2 - |\mathbb{E} Z|^2 = \text{Var } \Re Z + \text{Var } \Im Z,$$

while the pseudovariance is

$$\mathbb{V} Z = \mathbb{E} (Z - \mathbb{E} Z)^2 = \mathbb{E} Z^2 - (\mathbb{E} Z)^2 = \text{Var } \Re Z - \text{Var } \Im Z + 2i \text{Cov}(\Re Z, \Im Z).$$

Of course, these are equal for real variables.

2.1 The diameter of a complex random variable

Let X be an a.s. bounded real random variable on $\mathbf{P} = (\Omega, \mathcal{F}, P)$. As usual, define the *essential supremum* of X as

$$\text{ess sup } X = \sup \{x \in \mathbb{R} \mid P(X > x) > 0\}.$$

If $|X| \leq c$ a.s., it is well-known that $\text{ess sup } X = -c + \lim_{r \rightarrow \infty} (\mathbb{E}((X + c)^r))^{1/r}$. If Z is an a.s. bounded complex random variable on \mathbf{P} then we define the *diameter* of Z to be

$$\text{diam } Z = \text{ess sup } |Z - Z'|, \quad \text{where } Z' \text{ is an independent copy of } Z. \quad (2.1)$$

The probability in (2.1) is interpreted in the product space $\mathbf{P} \otimes \mathbf{P}$ in the standard fashion. We will also use an equivalent definition that does not use the product space. Given an angle θ , the extent of Z in the θ direction is $\Re(e^{-i\theta} Z)$, so we can alternatively define

$$\text{diam } Z = \sup_{\theta \in (-\pi, \pi]} (\text{ess sup}(\Re(e^{-i\theta} Z)) + \text{ess sup}(-\Re(e^{-i\theta} Z))). \quad (2.2)$$

Remark 2.1. To see that (2.1) and (2.2) are equivalent, suppose first that $\text{diam } Z > d + \varepsilon$ according to (2.1), for some $\varepsilon > 0$. Assuming that $|Z| \leq c$ a.s., cover the disk $\{z \mid |z| \leq c\}$ by finitely many open disks of radius $\varepsilon/4$. If for each pair D, D' of such disks whose centres are at least $d + \varepsilon/2$ apart we have $P(Z \in D, Z' \in D') = 0$, then $\text{ess sup } |Z - Z'| \leq d + \varepsilon$, a contradiction. So choose two of the disks, D, D' , with centres at least $d + \varepsilon/2$ apart, such that $P(Z \in D, Z' \in D') = P(Z \in D) P(Z' \in D') > 0$. Taking θ to be the direction from the centre of D to the centre of D' , we find that $\text{diam } Z \geq d$ according to (2.2). Conversely, if there is θ such that the argument of the \sup_θ in (2.2) is greater than d , there are half-planes more than d apart in each of which Z has nonzero probability, proving that $\text{diam } Z > d$ according to (2.1).

The basic properties of the diameter of a complex random variable are given by the following lemma.

Lemma 2.2. *Let Z be an a.s. bounded complex random variable on \mathbf{P} . Then,*

- (a) $\text{diam } Z = 0$ iff Z is a.s. constant.
- (b) $\text{diam } (aZ + b) = |a| \text{diam } Z$ for any $a, b \in \mathbb{C}$.
- (c) $\text{diam } (Z + W) \leq \text{diam } Z + \text{diam } W$ for any a.s. bounded complex random variable W on \mathbf{P} .
- (d) $\text{diam } \Re Z \leq \text{diam } Z \leq 2 \text{ess sup } |Z|$.
- (e) $\text{diam } (\Re Z)^2 \leq \text{diam } Z^2 \leq 2 \text{ess sup } |Z| \cdot \text{diam } Z$.
- (f) $|Z - \mathbb{E} Z| \leq \text{diam } Z$ a.s.
- (g) There exists $a \in \mathbb{C}$ such that $|Z - a| \leq \frac{1}{\sqrt{3}} \text{diam } Z$ a.s.
- (h) $\text{Var } Z = \mathbb{E} |Z - \mathbb{E} Z|^2 \leq \frac{1}{3} (\text{diam } Z)^2$ and $|\mathbb{V} Z| = |\mathbb{E} (Z - \mathbb{E} Z)^2| \leq \frac{1}{4} (\text{diam } Z)^2$.

Proof. Claims (a), (b) follow immediately from definition (2.1). We get claim (c) from definition (2.2) and the fact that $\text{ess sup } (X + Y) \leq \text{ess sup } X + \text{ess sup } Y$ for any a.s. bounded real random variables X, Y on \mathbf{P} .

Let Z' be an independent copy of Z . We note then (almost surely) that

$$\begin{aligned}
|\Re Z - \Re Z'| &\leq |Z - Z'| \leq \text{ess sup } |Z - Z'| = \text{diam } Z, \\
|(\Re Z)^2 - (\Re Z')^2| &= |\Re(Z - Z')| \cdot |\Re(Z + Z')| \leq |Z - Z'| \cdot |Z + Z'| \\
&\leq \text{ess sup } |Z^2 - (Z')^2| = \text{diam } Z^2. \\
|Z - Z'| &\leq \text{ess sup } |Z| + \text{ess sup } |Z'| = 2 \text{ess sup } |Z|. \\
|Z^2 - (Z')^2| &\leq \text{ess sup } |Z + Z'| \cdot \text{ess sup } |Z - Z'| \leq 2 \text{ess sup } |Z| \cdot \text{diam } Z.
\end{aligned}$$

Due to definition (2.1), claims (d) and (e) follow.

Using definition (2.2), the fact that $|X - EX| \leq \text{ess sup}(X) - \text{ess sup}(-X)$ a.s. for any a.s. bounded real random variable X on \mathbf{P} and the equation

$$|Z - \mathbb{E} Z| = \sup_{\theta \in (-\pi, \pi]} |\Re(e^{-i\theta}(Z - \mathbb{E} Z))|,$$

we obtain claim (f).

Claim (g) follows from a standard result on convex sets, see [29, Thm. 12.3], for example. An equilateral triangle shows that that the constant cannot be reduced. To prove the first part of claim (h), we note that $\text{Var } Z = \text{Var}(Z - a) \leq \text{ess sup } |Z - a|^2$.

However, for any a.s. bounded real random variable X on \mathbf{P}

$$|X - \frac{1}{2}(\text{ess sup } X + \text{ess sup}(-X))| \leq \frac{1}{2}(\text{ess sup } X - \text{ess sup}(-X)) \quad \text{a.s.},$$

which implies $\text{Var } X \leq \frac{1}{4}(\text{diam } X)^2$. To prove the second part of claim (h), we note that

$$|\mathbb{E}(Z - \mathbb{E} Z)^2| \leq \mathbb{E}(\Re(e^{-i\theta}(Z - \mathbb{E} Z))^2) = \text{Var } \Re(e^{-i\theta} Z),$$

where $e^{i\theta} = \mathbb{E}(Z - \mathbb{E} Z)^2 / |\mathbb{E}(Z - \mathbb{E} Z)^2|$, and $\text{diam}(\Re(e^{-i\theta} Z)) \leq \text{diam } Z$ on account of claims (b) and (d). \square

We will also use a conditional version of the diameter. Let $\mathcal{G} \subseteq \mathcal{F}$ be a σ -field. For a real random variable X on $\mathbf{P} = (\Omega, \mathcal{F}, P)$ such that $|X| \leq c$ a.s., we can define the *conditional essential supremum* of X to be the \mathcal{G} -measurable function

$$\text{ess sup}(X | \mathcal{G}) = -c + \lim_{r \rightarrow \infty} (\mathbb{E}((X + c)^r | \mathcal{G}))^{1/r}. \quad (2.3)$$

Alternative equivalent definitions and many properties of the conditional essential supremum are given in [1]. Informally, $\text{ess sup}(X | \mathcal{G})$ is the least \mathcal{G} -measurable function $G : \Omega \rightarrow \mathbb{R}$ such that $X \leq G$ a.s. Now we can extend (2.2) to define the *conditional diameter*:

$$\text{diam}(Z | \mathcal{G}) = \sup_{\theta \in (-\pi, \pi]} (\text{ess sup}(\Re(e^{-i\theta} Z) | \mathcal{G}) + \text{ess sup}(-\Re(e^{-i\theta} Z) | \mathcal{G})). \quad (2.4)$$

Note that $\text{diam}(Z | \mathcal{G})$ is a function from Ω to \mathbb{R}_+ . For any $\omega \in \Omega$, the argument of the \sup_θ in (2.4) is a continuous function of θ (since Z is a.s. bounded), so the supremum over θ is the same if restricted to a dense countable subset of $(-\pi, \pi]$. This proves that $\text{diam}(Z | \mathcal{G})$ is \mathcal{G} -measurable.

If Z is real, we can restrict (2.4) to $\theta = 0$ and then $\text{diam}(Z | \mathcal{G})$ is the same as the conditional range defined by McDiarmid [30, Sec. 3.4].

Now let $P_{Z|\mathcal{G}} : \mathcal{B}(\mathbb{C}) \times \Omega \rightarrow [0, 1]$ be a regular conditional distribution for Z given \mathcal{G} , where $\mathcal{B}(\mathbb{C})$ is the Borel field of \mathbb{C} . That is, for each $\omega \in \Omega$, $P_{Z|\mathcal{G}}(\cdot, \omega)$ is a probability measure on $\mathcal{B}(\mathbb{C})$, and for each $A \in \mathcal{B}(\mathbb{C})$, $P_{Z|\mathcal{G}}(A, \cdot)$ is \mathcal{G} -measurable and $P_{Z|\mathcal{G}}(A, \cdot) = P(Z^{-1}(A) | \mathcal{G})$ a.s. For the existence of $P_{Z|\mathcal{G}}$ and basic theory, see [26, Chap. 6].

For each $\omega \in \Omega$, let $K_\omega(Z | \mathcal{G})$ be the class of random variables from Ω to \mathbb{C} that induce the distribution $P_{Z|\mathcal{G}}(\cdot, \omega)$ on $\mathcal{B}(\mathbb{C})$. The most important property of $K_\omega(Z | \mathcal{G})$ is:

Lemma 2.3. *Let $\mathcal{G} \subseteq \mathcal{F}$ be a σ -field and Z be an a.s. bounded complex random variable on \mathbf{P} . Let Z_ω be an arbitrary member of $K_\omega(Z | \mathcal{G})$ for each $\omega \in \Omega$. Let W be a \mathcal{G} -measurable random variable on \mathbf{P} , and let $\phi : \mathbb{C} \times W(\Omega) \rightarrow \mathbb{C}$ be a measurable function such that $\mathbb{E}|\phi(Z, W)| < \infty$. Then, for almost all $\omega \in \Omega$,*

$$\mathbb{E}(\phi(Z, W) | \mathcal{G})(\omega) = \mathbb{E} \phi(Z_\omega, W), \quad (2.5)$$

and moreover $\phi(Z_\omega, W) \in K_\omega(\phi(Z, W) | \mathcal{G})$. Also the random variable $\omega \mapsto \mathbb{E} \phi(Z_\omega, W)$ is \mathcal{G} -measurable. Consequently, for almost all $\omega \in \Omega$,

$$\begin{aligned} \text{ess sup} (|\phi(Z, W)| | \mathcal{G})(\omega) &= \text{ess sup} |\phi(Z_\omega, W)|, \\ \text{diam}(\phi(Z, W) | \mathcal{G})(\omega) &= \text{diam} \phi(Z_\omega, W). \end{aligned} \quad (2.6)$$

Proof. Equation (2.5) is Theorem 6.4 in [26]. By applying it to functions of the form $\mathbf{1}_A(\phi(\cdot, \cdot))$ for each $A \in \mathcal{B}(\mathbb{C})$, we find that $\phi(Z_\omega, W) \in K_\omega(\phi(Z, W) | \mathcal{G})$. The \mathcal{G} -measurability of $\omega \mapsto \mathbb{E} \phi(Z_\omega, W)$ follows from the \mathcal{G} -measurability of the left side of (2.5). Equation (2.6) follows from (2.5) on account of (2.3) and (2.4). Note also that the \mathcal{G} -measurability of the conditional essential supremum and the conditional diameter is just a special case of this. \square

We now list a number of properties of the conditional diameter that we will need.

Lemma 2.4. *Let $\mathcal{G} \subseteq \mathcal{F}$ be a σ -field and Z be an a.s. bounded complex random variable on \mathbf{P} . Then,*

- (a) $\text{diam}(\Re Z | \mathcal{G}) \leq \text{diam}(Z | \mathcal{G})$ a.s.
- (b) $\text{diam}(Z | \mathcal{G}) \leq 2 \text{ess sup}(|Z| | \mathcal{G})$ a.s.
- (c) $\text{diam}(Z^2 | \mathcal{G}) \leq 2 \text{ess sup}(|Z| | \mathcal{G}) \text{diam}(Z | \mathcal{G})$ a.s.
- (d) $|Z - \mathbb{E}(Z | \mathcal{G})| \leq \text{diam}(Z | \mathcal{G})$ a.s.
- (e) If the σ -field $\mathcal{H} \subseteq \mathcal{F}$ is independent of \mathcal{G} , then $\text{diam} \mathbb{E}(Z | \mathcal{H}) \leq \mathbb{E} \text{diam}(Z | \mathcal{G})$ a.s.
- (f) $|\mathbb{E}((Z - \mathbb{E}(Z | \mathcal{G}))(W - \mathbb{E}(W | \mathcal{G})) | \mathcal{G})| \leq \frac{1}{3} \text{diam}(Z | \mathcal{G}) \cdot \text{diam}(W | \mathcal{G})$ a.s. for any a.s. bounded complex random variable W on \mathbf{P} .

(g) If U and W are \mathcal{G} -measurable, then

$$\text{diam}(UZ + W \mid \mathcal{G}) = |U| \text{diam}(Z \mid \mathcal{G}).$$

Proof. Let Z_ω be an arbitrary member of $K_\omega(Z \mid \mathcal{G})$ for each $\omega \in \Omega$. By Lemma 2.3, we have that, for almost all $\omega \in \Omega$,

$$\begin{aligned} \mathbb{E}(Z \mid \mathcal{G})(\omega) &= \mathbb{E} Z_\omega, \\ \text{ess sup}(|Z| \mid \mathcal{G})(\omega) &= \text{ess sup} |Z_\omega|, \\ \text{diam}(\Re Z \mid \mathcal{G})(\omega) &= \text{diam} \Re Z_\omega, \\ \text{diam}(Z \mid \mathcal{G})(\omega) &= \text{diam} Z_\omega, \\ \text{diam}(Z^2 \mid \mathcal{G})(\omega) &= \text{diam} Z_\omega^2, \\ \text{ess sup}(|Z - \mathbb{E}(Z \mid \mathcal{G})| \mid \mathcal{G})(\omega) &= \text{ess sup} |Z_\omega - \mathbb{E} Z_\omega|, \\ \mathbb{E}(|Z - \mathbb{E}(Z \mid \mathcal{G})|^2 \mid \mathcal{G})(\omega) &= \mathbb{E} |Z_\omega - \mathbb{E} Z_\omega|^2 = \text{Var} Z_\omega. \end{aligned}$$

Due to Lemma 2.2(d, e), claims (a)–(c) follow.

In order to prove claims (d) and (e), we recall from [1, Prop. 2.6] that for a bounded real random variable X ,

$$X \leq \text{ess sup}(X \mid \mathcal{G}) \quad \text{a.s.} \quad (2.7)$$

Therefore,

$$|Z - \mathbb{E}(Z \mid \mathcal{G})| \leq \text{ess sup}(|Z - \mathbb{E}(Z \mid \mathcal{G})| \mid \mathcal{G}) \quad \text{a.s.}$$

and we get claim (d) from Lemma 2.2(f).

Using (2.7) and the independence of \mathcal{G} and \mathcal{H} ,

$$\mathbb{E}(X \mid \mathcal{H}) \leq \mathbb{E}(\text{ess sup}(X \mid \mathcal{G}) \mid \mathcal{H}) = \mathbb{E} \text{ess sup}(X \mid \mathcal{G}) \quad \text{a.s.}$$

for a bounded real random variable X . Applying this to the definition (2.4) with $X = \Re(e^{-i\theta} Z)$ and $X = -\Re(e^{-i\theta} Z)$, claim (e) follows.

Claim (f) is due to Lemma 2.2(h) and the conditional Cauchy-Schwartz inequality

$$\begin{aligned} &|\mathbb{E}((Z - \mathbb{E}(Z \mid \mathcal{G}))(W - \mathbb{E}(W \mid \mathcal{G})) \mid \mathcal{G})| \\ &\leq \sqrt{\mathbb{E}(|Z - \mathbb{E}(Z \mid \mathcal{G})|^2 \mid \mathcal{G})} \sqrt{\mathbb{E}(|W - \mathbb{E}(W \mid \mathcal{G})|^2 \mid \mathcal{G})}. \end{aligned}$$

To prove claim (g), note that the properties of the conditional essential supremum imply

$$\text{ess sup}(\Re(e^{-i\theta}(U + WZ)) \mid \mathcal{G}) = \Re(e^{-i\theta}U) + |W| \text{ess sup}(\Re(e^{-i\theta + i \arg(W)} Z) \mid \mathcal{G}),$$

then apply the definition of the conditional diameter. \square

In [23] we proved the following generalization of a bound of Hoeffding [17].

Lemma 2.5. *If Z is an a.s. bounded complex random variable on \mathbf{P} , then*

$$|\mathbb{E} e^{Z - \mathbb{E} Z} - 1| \leq e^{\frac{1}{8} \text{diam}(Z)^2} - 1.$$

Corollary 2.6. *Let Z be an a.s. bounded complex random variable on \mathbf{P} and let $\mathcal{G} \subseteq \mathcal{F}$ be a σ -field. Then we have*

$$|\mathbb{E}(e^{Z - \mathbb{E}(Z|\mathcal{G})} | \mathcal{G}) - 1| \leq e^{\frac{1}{8} \text{diam}(Z|\mathcal{G})^2} - 1 \text{ a.s.}$$

Proof. It suffices to apply the lemma to arbitrary random variables $Z_\omega \in K_\omega(Z | \mathcal{G})$, with the help of (2.6). \square

2.2 First order approach

A sequence $\mathcal{F} = \mathcal{F}_0, \dots, \mathcal{F}_n$ of σ -subfields of \mathcal{F} is a *filter* if $\mathcal{F}_0 \subseteq \dots \subseteq \mathcal{F}_n$. A sequence Z_0, \dots, Z_n of random variables on $\mathbf{P} = (\Omega, \mathcal{F}, P)$ is a *martingale with respect to \mathcal{F}* if

- (i) Z_j is \mathcal{F}_j -measurable and has finite expectation, for $0 \leq j \leq n$;
- (ii) $\mathbb{E}(Z_j | \mathcal{F}_{j-1}) = Z_{j-1}$ for $1 \leq j \leq n$.

Note that, up to almost-sure equality, the martingale is determined by Z_n and \mathcal{F} , namely $Z_j = \mathbb{E}(Z_n | \mathcal{F}_j)$ a.s. for each j .

If Z is a random variable on \mathbf{P} and $0 \leq j \leq n$, we use the following notations for statistics conditional on \mathcal{F}_j :

$$\begin{aligned} \mathbb{E}_j Z &= \mathbb{E}(Z | \mathcal{F}_j), \\ \mathbb{V}_j Z &= \mathbb{E}((Z - \mathbb{E}_j(Z))^2 | \mathcal{F}_j) = \mathbb{E}_j Z^2 - (\mathbb{E}_j Z)^2, \\ \text{diam}_j Z &= \text{diam}(Z | \mathcal{F}_j). \end{aligned}$$

If $\mathcal{F}_0 = \{\emptyset, \Omega\}$, which we not assume unless it is stated explicitly, $\mathbb{E}_0 Z$, $\mathbb{V}_0 Z$ and $\text{diam}_0 Z$ equal the unconditional versions $\mathbb{E} Z$, $\mathbb{V} Z$ and $\text{diam} Z$, respectively.

An extremely large literature concerns concentration of martingales derived from restrictions on the differences $Z_j - Z_{j-1}$, but most of it considers only real martingales and can't be assumed to hold for complex martingales. The fact that the real and imaginary parts of a complex martingale are real martingales can often be applied, but at the cost of weaker bounds. In any case, our aim is for estimates of the exponential rather than for concentration. Here again, the non-convexity of the exponential function in the

complex plane often means that theorems and proofs for the real case do not carry over immediately to the complex case.

Theorem 2.7. *Let $\mathbf{Z} = Z_0, Z_1, \dots, Z_n$ be an a.s. bounded complex-valued martingale with respect to a filter $\mathcal{F}_0, \dots, \mathcal{F}_n$. Define*

$$R_k = \text{diam}_{k-1} Z_k \quad (2.8)$$

for $1 \leq k \leq n$. Then

$$\mathbb{E}_0 e^{Z_n} = e^{Z_0} (1 + K(\mathbf{Z})),$$

where $K(\mathbf{Z})$ is an \mathcal{F}_0 -measurable random variable with

$$|K(\mathbf{Z})| \leq \text{ess sup} \left(e^{\frac{1}{8} \sum_{k=1}^n R_k^2} \mid \mathcal{F}_0 \right) - 1 \quad \text{a.s.}$$

Proof. Since $\mathbb{E}_{k-1} Z_k = Z_{k-1}$, we have for $1 \leq k \leq n$ that

$$\begin{aligned} \mathbb{E}_{k-1} e^{Z_k} &= \mathbb{E}_{k-1} e^{Z_{k-1} + (Z_k - Z_{k-1})} \\ &= e^{Z_{k-1}} (1 + \mathbb{E}_{k-1}(e^{Z_k - Z_{k-1}} - 1)) \\ &= e^{Z_{k-1}} + U_k e^{Z_{k-1}} \end{aligned} \quad (2.9)$$

for some \mathcal{F}_{k-1} -measurable U_k such that $|U_k| \leq e^{R_k^2/8} - 1$ a.s., by Corollary 2.6.

Now recall that $|e^z| = e^{\Re z}$ for all z and note that $\Re Z_0, \dots, \Re Z_n$ is also a martingale satisfying the conditions of the theorem on account of Lemma 2.4(a). Therefore we similarly have that

$$\mathbb{E}_{k-1} |e^{Z_k}| = |e^{Z_{k-1}}| + U'_k |e^{Z_{k-1}}| = (1 + U'_k) |e^{Z_{k-1}}| \quad (2.10)$$

for some \mathcal{F}_{k-1} -measurable U'_k such that $|U'_k| \leq e^{R_k^2/8} - 1$ a.s. Now we can prove by backwards induction on k that for $0 \leq k \leq n$,

$$\mathbb{E}_k e^{Z_n} = e^{Z_k} + W_k e^{Z_k}, \quad (2.11)$$

where W_k is \mathcal{F}_k -measurable and $|W_k| \leq \text{ess sup} \left(e^{\frac{1}{8} \sum_{j=k+1}^n R_j^2} \mid \mathcal{F}_k \right) - 1$ a.s. Obviously (2.11) is true for $k = n$. Now observe from (2.9) and (2.11) that $\mathbb{E}_{k-1} Z_n = e^{Z_{k-1}} + U_k e^{Z_{k-1}} + \mathbb{E}_{k-1}(W_k e^{Z_k})$ and note that $|\mathbb{E}_{k-1}(W_k e^{Z_k})| \leq \text{ess sup}(|W_k| \mid \mathcal{F}_{k-1}) \mathbb{E}_{k-1} |e^{Z_k}|$ a.s. Applying (2.10) to the last term and combining the error terms using Lemma 6.2, we obtain (2.11) for $k - 1$. The case $k = 0$ gives the theorem. \square

2.3 Second order approach

In the following we need two technical bounds that are in the Appendix, Lemma 6.1. We also use the following elementary lemma.

Lemma 2.8. *Let $\mathbf{Z} = Z_0, Z_1, \dots, Z_n$ be a bounded complex-valued martingale with respect to a filter $\mathcal{F}_0, \dots, \mathcal{F}_n$. Then*

$$\mathbb{E}_k(Z_n - Z_k)^2 = \sum_{j=k+1}^n \mathbb{E}_k(Z_j - Z_{j-1})^2$$

for $0 \leq k \leq n$.

Proof. For $0 \leq k \leq j \leq \ell \leq n$,

$$\mathbb{E}_k(Z_\ell - Z_j)^2 = \mathbb{E}_k(\mathbb{E}_j(Z_\ell - Z_j)^2) = \mathbb{E}_k(\mathbb{E}_j Z_\ell^2 - Z_j^2) = \mathbb{E}_k(Z_\ell^2 - Z_j^2).$$

Therefore,

$$\begin{aligned} \mathbb{E}_k(Z_n - Z_k)^2 &= \mathbb{E}_k(Z_n^2 - Z_k^2) \\ &= \mathbb{E}_k((Z_n^2 - Z_{n-1}^2) + (Z_{n-1}^2 - Z_{n-2}^2) + \dots + (Z_{k+1}^2 - Z_k^2)) \\ &= \sum_{j=k+1}^n \mathbb{E}_k(Z_j - Z_{j-1})^2. \end{aligned} \quad \square$$

Theorem 2.9. *Let $\mathbf{Z} = Z_0, Z_1, \dots, Z_n$ be a bounded complex-valued martingale with respect to a filter $\mathcal{F}_0, \dots, \mathcal{F}_n$. For $1 \leq k \leq n$, define*

$$R_k = \text{diam}_{k-1} Z_k, \tag{2.12a}$$

$$Q_k = \max\{\text{diam}_{k-1} \mathbb{E}_k(Z_n - Z_k)^2, \text{diam}_{k-1} \mathbb{E}_k(\Re Z_n - \Re Z_k)^2\}. \tag{2.12b}$$

Then

$$\begin{aligned} \mathbb{E}_0 e^{Z_n} &= e^{Z_0 + \frac{1}{2} \mathbb{V}_0 Z_n} + L(\mathbf{Z}) e^{\Re Z_0 + \frac{1}{2} \mathbb{V}_0(\Re Z_n)} \\ &= e^{Z_0 + \frac{1}{2} \mathbb{V}_0 Z_n} (1 + L'(\mathbf{Z}) e^{\frac{1}{2} \mathbb{V}_0(\Im Z_n)}), \end{aligned}$$

where $L(\mathbf{Z})$, $L'(\mathbf{Z})$ are \mathcal{F}_0 -measurable random variables with

$$|L(\mathbf{Z})| = |L'(\mathbf{Z})| \leq \text{ess sup} \left(\exp \left(\sum_{k=1}^n \left(\frac{1}{6} R_k^3 + \frac{1}{6} R_k Q_k + \frac{5}{8} R_k^4 + \frac{5}{32} Q_k^2 \right) \right) \middle| \mathcal{F}_0 \right) - 1 \quad a.s.$$

Proof. All equalities and inequalities in the proof should be taken “almost surely”. For $1 \leq k \leq n$ we have, using Lemma 2.8,

$$\begin{aligned}
\mathbb{E}_{k-1} e^{Z_k + \frac{1}{2} \mathbb{V}_k Z_n} &= e^{Z_{k-1} + \frac{1}{2} \mathbb{V}_{k-1} Z_n} \\
&\quad + e^{Z_{k-1} + \frac{1}{2} \mathbb{E}_{k-1}(Z_n - Z_k)^2} \mathbb{E}_{k-1}((e^{A_k} - e^{A_k^2/2} - A_k)e^{B_k}) \\
&\quad + e^{Z_{k-1} + \frac{1}{2} \mathbb{E}_{k-1}(Z_n - Z_k)^2} \mathbb{E}_{k-1}(A_k(e^{B_k} - B_k - 1)) \\
&\quad + e^{Z_{k-1} + \frac{1}{2} \mathbb{E}_{k-1}(Z_n - Z_k)^2} \mathbb{E}_{k-1}(A_k + A_k B_k) \\
&\quad + e^{Z_{k-1} + \frac{1}{2} \mathbb{V}_{k-1} Z_n} \mathbb{E}_{k-1}(e^{C_k} - 1),
\end{aligned} \tag{2.13}$$

where

$$\begin{aligned}
A_k &= Z_k - Z_{k-1}, \\
B_k &= \frac{1}{2} \mathbb{E}_k(Z_n - Z_k)^2 - \frac{1}{2} \mathbb{E}_{k-1}(Z_n - Z_k)^2 = \frac{1}{2} \mathbb{V}_k Z_n - \frac{1}{2} \mathbb{E}_{k-1} \mathbb{V}_k Z_n, \\
C_k &= \frac{1}{2}(Z_k - Z_{k-1})^2 + \frac{1}{2} \mathbb{V}_k Z_n - \frac{1}{2} \mathbb{V}_{k-1} Z_n \\
&= B_k + \frac{1}{2} A_k^2 - \frac{1}{2} \mathbb{E}_{k-1} A_k^2.
\end{aligned}$$

Note that

$$\mathbb{E}_{k-1} A_k = \mathbb{E}_{k-1} B_k = \mathbb{E}_{k-1} C_k = 0.$$

Therefore, by the conditions of the theorem and Lemma 2.4(b,d),

$$\begin{aligned}
|A_k| &\leq R_k, \\
|B_k| &\leq \text{diam}_{k-1} B_k = \frac{1}{2} \text{diam}_{k-1} \mathbb{E}_k(Z_n - Z_k)^2 \leq \frac{1}{2} Q_k \text{ and} \\
|C_k| &\leq \text{diam}_{k-1} C_k \leq \frac{1}{2} \text{diam}_{k-1} (Z_k - Z_{k-1})^2 + \frac{1}{2} \text{diam}_{k-1} \mathbb{E}_k(Z_n - Z_k)^2 \\
&\leq \text{ess sup} (|Z_k - Z_{k-1}|^2 \mid \mathcal{F}_{k-1}) + \frac{1}{2} Q_k \leq R_k^2 + \frac{1}{2} Q_k
\end{aligned}$$

By Corollary 2.6,

$$|\mathbb{E}_{k-1}(e^{C_k} - 1)| \leq e^{\frac{1}{8}(R_k^2 + Q_k/2)^2} - 1 \leq e^{\frac{1}{4}R_k^4 + \frac{1}{16}Q_k^2} - 1.$$

Using Lemma 6.1 and Corollary 2.6 with the triangle inequality, we get that

$$\begin{aligned}
|\mathbb{E}_{k-1}((e^{A_k} - e^{A_k^2/2} - A_k)e^{B_k})| &\leq (e^{\frac{1}{6}R_k^3 + \frac{1}{8}R_k^4} - 1) \mathbb{E}_{k-1}(|e^{B_k}|) \\
&= (e^{\frac{1}{6}R_k^3 + \frac{1}{8}R_k^4} - 1)(\mathbb{E}_{k-1}(e^{\Re B_k} - 1) + 1) \leq e^{\frac{1}{6}R_k^3 + \frac{1}{8}R_k^4 + \frac{1}{32}Q_k^2} - e^{\frac{1}{32}Q_k^2}, \\
|\mathbb{E}_{k-1}(A_k(e^{B_k} - B_k - 1))| &\leq e^{\frac{1}{32}Q_k^2} + e^{\frac{1}{6}R_k Q_k + \frac{1}{16}Q_k^2 + \frac{1}{4}R_k^4} - \frac{1}{6}R_k Q_k - 2.
\end{aligned}$$

By Lemma 2.4(f), we have

$$|\mathbb{E}_{k-1} A_k B_k| \leq \frac{1}{3} \text{diam}_{k-1} A_k \cdot \text{diam}_{k-1} B_k \leq \frac{1}{6} R_k Q_k.$$

Therefore, for each k , formula (2.13) gives

$$\begin{aligned}\mathbb{E}_{k-1} e^{Z_k + \frac{1}{2} \mathbb{V}_k Z_n} &= e^{Z_{k-1} + \frac{1}{2} \mathbb{V}_{k-1} Z_n} + L_k e^{Z_{k-1} + \frac{1}{2} \mathbb{V}_{k-1} Z_n} \\ &\quad + L'_k e^{Z_{k-1} + \frac{1}{2} \mathbb{E}_{k-1}(Z_n - Z_k)^2}\end{aligned}\tag{2.14}$$

for some \mathcal{F}_{k-1} -measurable random variables L_k and L'_k with

$$\begin{aligned}|L_k| &\leq e^{\frac{1}{4}R_k^4 + \frac{1}{16}Q_k^2} - 1, \\ |L'_k| &\leq e^{\frac{1}{6}R_k^3 + \frac{3}{8}R_k^4 + \frac{1}{6}R_k Q_k + \frac{3}{32}Q_k^2} - 1.\end{aligned}$$

Now consider the martingale X_0, \dots, X_n of the real parts of Z_0, \dots, Z_n . In order to bound the second and third terms of (2.14) we consider the absolute value

$$\begin{aligned}|e^{Z_{k-1} + \frac{1}{2} \mathbb{V}_{k-1} Z_n}| &= e^{\Re Z_{k-1} + \frac{1}{2} \mathbb{V}_{k-1} \Re Z_n - \frac{1}{2} \mathbb{V}_{k-1} \Im Z_n} \leq e^{X_{k-1} + \frac{1}{2} \mathbb{V}_{k-1} X_n}, \\ |e^{Z_{k-1} + \frac{1}{2} \mathbb{E}_{k-1}(Z_n - Z_k)^2}| &= e^{\Re Z_{k-1} + \frac{1}{2} \mathbb{E}_{k-1}(\Re Z_n - \Re Z_k)^2 - \frac{1}{2} \mathbb{E}_{k-1}(\Im Z_n - \Im Z_k)^2} \\ &\leq e^{X_{k-1} + \frac{1}{2} \mathbb{E}_{k-1}(X_n - X_k)^2} \\ &= e^{X_{k-1} + \frac{1}{2} \mathbb{V}_{k-1} X_n - \frac{1}{2} \mathbb{E}_{k-1}(X_k - X_{k-1})^2} \\ &\leq e^{X_{k-1} + \frac{1}{2} \mathbb{V}_{k-1} X_n}.\end{aligned}\tag{2.15}$$

Due to Lemma 2.4(a), this martingale also satisfies conditions (2.12). Therefore, by the same reasoning as before and using the inequality

$$e^{X_{k-1} + \frac{1}{2} \mathbb{E}_{k-1}(X_n - X_k)^2} \leq e^{X_{k-1} + \frac{1}{2} \mathbb{V}_{k-1} X_n},$$

we get that

$$\mathbb{E}_{k-1} e^{X_k + \frac{1}{2} \mathbb{V}_k X_n} = e^{X_{k-1} + \frac{1}{2} \mathbb{V}_{k-1} X_n} + L''_k e^{X_{k-1} + \frac{1}{2} \mathbb{V}_{k-1} X_n},\tag{2.16}$$

where $|L''_k| \leq e^{\frac{1}{4}R_k^4 + \frac{1}{16}Q_k^2} - 1 + e^{\frac{1}{6}R_k^3 + \frac{3}{8}R_k^4 + \frac{1}{6}R_k Q_k + \frac{3}{32}Q_k^2} - 1 \leq e^{\frac{1}{6}R_k^3 + \frac{5}{8}R_k^4 + \frac{1}{6}R_k Q_k + \frac{5}{32}Q_k^2} - 1$.

Using (2.14) and (2.16), we now prove by backwards induction on k that

$$\mathbb{E}_k e^{Z_n} = e^{Z_k + \frac{1}{2} \mathbb{V}_k Z_n} + M_k e^{X_k + \frac{1}{2} \mathbb{V}_k X_n},\tag{2.17}$$

where

$$|M_k| \leq \text{ess sup} \left(e^{\sum_{j=k+1}^n (\frac{1}{6}R_j^3 + \frac{5}{8}R_j^4 + \frac{1}{6}R_j Q_j + \frac{5}{32}Q_j^2)} \mid \mathcal{F}_k \right) - 1.$$

The claim is obviously true for $k = n$. To perform the induction step, take the expectation of (2.17) with respect to \mathcal{F}_{k-1} , using (2.14) and (2.15) for the first term on the right, and (2.16) to bound the second term on the right. Using the bound $e^{\frac{1}{6}R_k^3 + \frac{5}{8}R_k^4 + \frac{1}{6}R_k Q_k + \frac{5}{32}Q_k^2} - 1$ for both $|L_k| + |L'_k|$ and $|L''_k|$, we obtain (2.17) for $\mathbb{E}_{k-1} e^{Z_n}$ on combining the error terms using Lemma 6.2. After n steps we reach the expression for $\mathbb{E}_0 e^{Z_n}$ stated in the theorem. \square

Remark 2.10. Although the two options on the right side of (2.12b) are the same for real martingales, either one of them can be the largest for complex martingales. The case of a purely imaginary martingale shows that the first can be larger. To show that the second can be larger, consider independent variables X, Y , where $X \in \{1, e^{2\pi i/3}, e^{-2\pi i/3}\}$ with equal probabilities, and $Y \in \{0, 1\}$ with equal probabilities. Now consider the martingale Z_0, Z_1, Z_2 where $Z_2 = XY$, $Z_1 = \mathbb{E}(Z_2 \mid \mathcal{F}_1) = 0$ and $Z_0 = \mathbb{E}(Z_1 \mid \mathcal{F}_0) = 0$, where $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_1 = \sigma(Y)$. We find that $\mathbb{E}_1(Z_2 - Z_1)^2 = 0$ and $\mathbb{E}_1(\Re Z_2 - \Re Z_1)^2 \in \{\frac{1}{2}, 1\}$ with probabilities $\frac{2}{3}, \frac{1}{3}$ respectively. Therefore $\text{diam}_0(\mathbb{E}_1(Z_2 - Z_1)^2) = 0 < \frac{1}{2} = \text{diam}_0(\mathbb{E}_1(\Re Z_2 - \Re Z_1)^2)$.

3 Functions of independent random variables

In this section we apply our martingale theorems to the case of functions of independent random variables.

An important example of a martingale is made by the so-called *Doob martingale process*. Suppose X_1, X_2, \dots, X_n are random variables and $f(X_1, X_2, \dots, X_n)$ is a random variable of bounded expectation. Then we have the martingale $\{Z_j\}$ with respect to $\{\mathcal{F}_j\}$, where for each j , $\mathcal{F}_j = \sigma(X_1, \dots, X_j)$ and $Z_j = \mathbb{E}(f(X_1, X_2, \dots, X_n) \mid \mathcal{F}_j)$. In particular, $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $Z_0 = \mathbb{E}f(X_1, X_2, \dots, X_n)$. We will also use the σ -fields $\mathcal{F}^{(j)} = \sigma(X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_n)$ for $1 \leq j \leq n$.

In this section we use the Doob martingale to find bounds on $\mathbb{E}e^f$. We first need some preliminary lemmas in order to show that all assumptions of Theorems 2.7 and 2.9 are satisfied.

Lemma 3.1. *Suppose that X, Y are independent random variables on \mathbf{P} , and that g is a complex-valued function such that $g(X, Y)$ is bounded and measurable. Then,*

- (a) $\text{diam}(g(X, Y) \mid \sigma(X))(\omega) = \text{diam } g(X, Y(\omega))$ for almost all $\omega \in \Omega$.
- (b) $\text{diam}(g(X, Y) \mid \sigma(X)) \leq \sup_{x \in X(\Omega), y, y' \in Y(\Omega)} |g(x, y) - g(x, y')|$ a.s.
- (c) $\text{diam}(g(X, Y) - \mathbb{E}(g(X, Y) \mid \sigma(Y)) \mid \sigma(X))$
 $\leq \sup_{x, x' \in X(\Omega), y, y' \in Y(\Omega)} |g(x, y) - g(x', y) - g(x, y') + g(x', y')|$ a.s.

Proof. Since Y is by definition $\sigma(Y)$ -measurable, Lemma 2.3 tells us that $g(X(\omega), Y) \in K_\omega(g(X, Y) \mid \sigma(X))$. Claim (a) is thus just the definition of the conditional diameter. Similarly, $g(X(\omega), Y) \in K_\omega(g(X, Y) \mid \sigma(X))$, which gives claim (b) when we apply the definition of $\text{diam}(g(X(\omega), Y))$.

For claim (c), note that for almost all $\omega \in \Omega$, $\mathbb{E}(g(X, Y) \mid \sigma(Y))(\omega) = \mathbb{E}g(X, Y(\omega))$. Therefore, applying claim (b),

$$\begin{aligned} \text{diam}(g(X, Y) - \mathbb{E}(g(X, Y) \mid \sigma(Y)) \mid \sigma(X)) \\ \leq \sup_{x \in X(\Omega), y, y' \in Y(\Omega)} |g(x, y) - \mathbb{E}(X, y) - g(x, y') + \mathbb{E}(X, y')| \\ \leq \sup_{x, x' \in X(\Omega), y, y' \in Y(\Omega)} |g(x, y) - g(x', y) - g(x, y') + g(x', y')| \end{aligned}$$

since $|\mathbb{E}U| \leq \sup|U|$ for any complex random variable. \square

We will deal with functions with additional arguments. For these purposes we state the following corollary of Lemma 2.4 and Lemma 3.1.

Corollary 3.2. *Suppose that W, X, Y are independent random variables on \mathbf{P} , and that h is a complex-valued function such that $h(W, X, Y)$ is bounded and measurable. Then, a.s.,*

$$\begin{aligned} (a) \quad & \text{diam}(\mathbb{E}(h(W, X, Y) \mid \sigma(W, X)) \mid \sigma(W)) \leq \mathbb{E}(\text{diam}(h(W, X, Y) \mid \sigma(W, Y)) \mid \sigma(W)). \\ (b) \quad & \text{diam}(h(W, X, Y) - \mathbb{E}(h(W, X, Y) \mid \sigma(W, Y)) \mid \sigma(W, X)) \\ & \leq \sup_{w \in W(\Omega), x, x' \in X(\Omega), y, y' \in Y(\Omega)} |h(w, x, y) - h(w, x', y) - h(w, x, y') + h(w, x', y')|. \end{aligned}$$

Proof. Using Lemma 3.1(a), we note that for almost all $\omega \in \Omega$,

$$\begin{aligned} \mathbb{E}(h(W, X, Y) \mid \sigma(W, X))(\omega) &= \mathbb{E}(h(W(\omega), X(\omega), Y)) \\ &= \mathbb{E}(h(W(\omega), X, Y) \mid \sigma(X))(\omega), \\ \text{diam}(h(W, X, Y) \mid \sigma(W, X))(\omega) &= \text{diam}(h(W(\omega), X(\omega), Y)) \\ &= \text{diam}(h(W(\omega), X, Y) \mid \sigma(X))(\omega), \end{aligned}$$

and the same with X and Y interchanged, so we can apply Lemma 2.4(e) and Lemma 3.1(c) to random variables given by the two-argument functions $g_\omega(X, Y) = h(W(\omega), X, Y)$ to obtain both claims. \square

3.1 Estimating the exponential

In this section we state our main results when applied to the case of complex functions of independent random variables. Let $\mathbf{P} = (\Omega, \mathcal{F}, P)$ be a probability space. Let $\mathbf{S} = S_1 \times \cdots \times S_n$ be any n -dimensional domain and consider a function $F : \mathbf{S} \rightarrow \mathbb{C}$. For $1 \leq k \leq n$, define

$$\alpha_k(F, \mathbf{S}) = \sup |F(\mathbf{x}^k) - F(\mathbf{x})|, \quad (3.1)$$

where the supremum is over $\mathbf{x}, \mathbf{x}^k \in \mathbf{S}$ that differ only in the k -th coordinate. Similarly, for $j \neq k$, define

$$\Delta_{jk}(F, \mathbf{S}) = \sup |F(\mathbf{x}) - F(\mathbf{x}^j) - F(\mathbf{x}^k) + F(\mathbf{x}^{jk})|, \quad (3.2)$$

where the supremum is over $\mathbf{x}, \mathbf{x}^k, \mathbf{x}^j, \mathbf{x}^{jk} \in \mathbf{S}$ such that \mathbf{x}, \mathbf{x}^k differ only in the k -th component, \mathbf{x}, \mathbf{x}^j differ only in the j -th component, $\mathbf{x}^j, \mathbf{x}^{jk}$ only in the k -th component, and $\mathbf{x}^k, \mathbf{x}^{jk}$ only in the j -th component. We also define the column vector $\boldsymbol{\alpha}(F, \mathbf{S}) = (\alpha_1(F, \mathbf{S}), \dots, \alpha_n(F, \mathbf{S}))^T$ and the matrix $\Delta(F, \mathbf{S}) = (\Delta_{jk}(F, \mathbf{S}))$ with zero diagonal.

Theorem 3.3. *Let $\mathbf{X} = (X_1, \dots, X_n)$ be a random vector on \mathbf{P} with independent components, and let $f : \mathbf{X}(\Omega) \rightarrow \mathbb{C}$ be a measurable function. Let $\boldsymbol{\alpha} = \boldsymbol{\alpha}(f, \mathbf{X}(\Omega))$ and $\Delta = \Delta(f, \mathbf{X}(\Omega))$.*

(a) *We have*

$$\mathbb{E} e^{f(\mathbf{X})} = e^{\mathbb{E} f(\mathbf{X})} (1 + K), \quad (3.3)$$

where $K = K(f(\mathbf{X}))$ is a complex constant with $|K| \leq e^{\frac{1}{8} \boldsymbol{\alpha}^T \boldsymbol{\alpha}} - 1$.

(b) *We have*

$$\begin{aligned} \mathbb{E} e^{f(\mathbf{X})} &= e^{\mathbb{E} f(\mathbf{X}) + \frac{1}{2} \mathbb{V} f(\mathbf{X})} + L e^{\mathbb{E} \Re f(\mathbf{X}) + \frac{1}{2} \text{Var} \Re f(\mathbf{X})} \\ &= e^{\mathbb{E} f(\mathbf{X}) + \frac{1}{2} \mathbb{V} f(\mathbf{X})} (1 + L' e^{\frac{1}{2} \text{Var} \Im f(\mathbf{X})}), \end{aligned} \quad (3.4)$$

where $L = L(f(\mathbf{X}))$ and $L' = L'(f(\mathbf{X}))$ are complex constants with

$$|L| = |L'| \leq \exp \left(\frac{1}{6} \sum_{k=1}^n \alpha_k^3 + \frac{1}{6} \boldsymbol{\alpha}^T \Delta \boldsymbol{\alpha} + \frac{5}{8} \sum_{k=1}^n \alpha_k^4 + \frac{5}{16} \boldsymbol{\alpha}^T \Delta^2 \boldsymbol{\alpha} \right) - 1.$$

Proof. Consider the martingale $\{Z_k\}$ with respect to $\{\mathcal{F}_k\}$ obtained by the Doob martin-

gale process. For $1 \leq k \leq n$ we have

$$\begin{aligned}
\text{diam}_{k-1} Z_k &= \text{diam}(\mathbb{E}(f(\mathbf{X}) \mid \mathcal{F}_k) \mid \mathcal{F}_{k-1}) \\
&\leq \mathbb{E}(\text{diam}(f(\mathbf{X}) \mid \mathcal{F}^{(k)}) \mid \mathcal{F}_{k-1}), \quad \text{by Corollary 3.2(a)} \\
&\leq \text{ess sup}(\text{diam}(f(\mathbf{X}) \mid \mathcal{F}^{(k)})) \\
&\leq \sup_{\mathbf{x}, \mathbf{x}^k} |f(\mathbf{x}) - f(\mathbf{x}^k)|, \quad \text{by Lemma 3.1(b)} \\
&\leq \alpha_k, \quad \text{by assumption.}
\end{aligned}$$

Now formula (3.3) follows from Theorem 2.7.

We next consider $\mathbb{E}_k(Z_n - Z_k)^2$, which by Lemma 2.8 is equal to $\sum_{j=k+1}^n \mathbb{E}_k(Z_j - Z_{j-1})^2$.

$$\begin{aligned}
&\text{diam}_{k-1} \mathbb{E}_k(Z_j - Z_{j-1})^2 \\
&= \text{diam}(\mathbb{E}((Z_j - Z_{j-1})^2 \mid \mathcal{F}_k) \mid \mathcal{F}_{k-1}), \\
&\leq \mathbb{E}(\text{diam}((Z_j - Z_{j-1})^2 \mid \mathcal{F}^{(k)} \cap \mathcal{F}_j) \mid \mathcal{F}_{k-1}) \quad \text{by Corollary 3.2(a).} \quad (3.5) \\
&\leq 2 \text{ess sup} |Z_j - Z_{j-1}| \cdot \text{ess sup} \left| \text{diam}(Z_j - Z_{j-1} \mid \mathcal{F}^{(k)} \cap \mathcal{F}_j) \right|, \\
&\quad \text{by Lemma 2.4(c).}
\end{aligned}$$

By Lemma 2.4(d),

$$|Z_j - Z_{j-1}| \leq \text{diam}_{j-1} Z_j \leq \alpha_j. \quad (3.6)$$

Using Corollary 3.2(a,b), we find that

$$\begin{aligned}
\text{diam}(Z_j - Z_{j-1} \mid \mathcal{F}^{(k)} \cap \mathcal{F}_j) &= \text{diam}(\mathbb{E}_j(f(\mathbf{X}) - \mathbb{E}(f(\mathbf{X}) \mid \mathcal{F}^{(j)})) \mid \mathcal{F}^{(k)} \cap \mathcal{F}_j) \\
&\leq \mathbb{E}(\text{diam}(f(\mathbf{X}) - \mathbb{E}(f(\mathbf{X}) \mid \mathcal{F}^{(j)}) \mid \mathcal{F}^{(k)}) \mid \mathcal{F}^{(k)} \cap \mathcal{F}_j) \\
&\leq \text{ess sup} \text{diam}(f(\mathbf{X}) - \mathbb{E}(f(\mathbf{X}) \mid \mathcal{F}^{(j)}) \mid \mathcal{F}^{(k)}) \\
&\leq \sup |f(\mathbf{x}) - f(\mathbf{x}^j) - f(\mathbf{x}^k) + f(\mathbf{x}^{jk})| \\
&\leq \Delta_{jk}, \quad \text{by assumption,} \quad (3.7)
\end{aligned}$$

where the last supremum is over $\mathbf{x}, \mathbf{x}^k, \mathbf{x}^j, \mathbf{x}^{jk} \in \mathbf{X}(\Omega)$ such that \mathbf{x}, \mathbf{x}^k differ only in the k -th component, \mathbf{x}, \mathbf{x}^j differ only in the j -th component, $\mathbf{x}^j, \mathbf{x}^{jk}$ only in the k -th component, and $\mathbf{x}^k, \mathbf{x}^{jk}$ only in the j -th component. Combining (3.5)–(3.7), we obtain that

$$\text{diam}_{k-1} \mathbb{E}_k(Z_j - Z_{j-1})^2 \leq 2\alpha_j \Delta_{jk}.$$

The same bound holds for $\text{diam}_{k-1} \mathbb{E}_k(\Re Z_n - \Re Z_k)^2$, since the Doob martingale of $\Re f(\mathbf{X})$ also satisfies conditions (a) and (b) of the theorem.

Now we can apply Theorem 2.9 to obtain (3.3) with

$$|L(f, \mathbf{X})| \leq \exp \left(\frac{1}{6} \sum_{k=1}^n \alpha_k^3 + \frac{5}{8} \sum_{k=1}^n \alpha_k^4 + \frac{1}{3} \sum_{k=1}^n \sum_{j=k+1}^n \alpha_j \alpha_k \Delta_{jk} + \frac{5}{8} \sum_{k=1}^n \left(\sum_{j=k+1}^n \alpha_j \Delta_{jk} \right)^2 \right) - 1.$$

Since the matrix Δ is symmetric, the third term in the summation equals $\frac{1}{6} \boldsymbol{\alpha}^T \Delta \boldsymbol{\alpha}$.

The term $A = \sum_{k=1}^n \left(\sum_{j=k+1}^n \alpha_j \Delta_{jk} \right)^2$ depends on the order that the arguments of f are listed, but we can define the martingale using any order we wish. If we write $A = \sum_{j>k, \ell>k} \alpha_j \Delta_{jk} \Delta_{k\ell} \alpha_\ell$, then the version from the reverse order of the arguments is $A' = \sum_{j<k, \ell<k} \alpha_j \Delta_{jk} \Delta_{k\ell} \alpha_\ell$. Since A and A' provide disjoint sets of terms of $\boldsymbol{\alpha}^T \Delta^2 \boldsymbol{\alpha} = \sum_{j,k,\ell} \alpha_j \Delta_{jk} \Delta_{k\ell} \alpha_\ell$, at least one of them is bounded by $\frac{1}{2} \boldsymbol{\alpha}^T \Delta^2 \boldsymbol{\alpha}$. This completes the proof. \square

Remark 3.4. A result similar to Theorem 3.3 was proved by Catoni [10, 11] when the function f is real, and used to obtain concentration bounds of the form

$$P(f(\mathbf{X}) \geq \mathbb{E} f(\mathbf{X}) + t) \leq \exp \left(- \frac{t^2}{2(\text{Var } f(\mathbf{X}) + \eta t / \text{Var } f(\mathbf{X}))} \right),$$

where η is a certain constant depending on $\boldsymbol{\alpha}$ and Δ . We won't pursue that direction here since we are interested in the complex case which is required for our applications. The complex case has the added advantage that we can use it to estimate characteristic functions and not just Laplace transforms, with interesting consequences that include Berry–Esseen-type inequalities which we will explore in a further paper.

Another point to mention in comparison with Catoni's theorems is that he doesn't have fourth-order terms such as the term $\frac{5}{8} \sum_{k=1}^n \alpha_k^4$ in Theorem 3.3(b). Although those terms make the bound much larger for very large $\{\alpha_j\}$, in such extreme cases part (a) of the theorem generally gives a better result anyway. We have included fourth order terms in order to allow better constants on the third order terms.

Remark 3.5. The factor $e^{\frac{1}{2} \text{Var } \Im f(\mathbf{X})}$ appearing in the error term of (3.4) is of course redundant in the case that f is real. The following example shows that some such multiplier is required in the general complex case. Suppose that the components of $\mathbf{X} = (X_1, \dots, X_n)$ are iid random variables with mass $\frac{1}{2}$ at each of $\pm n^{-1/2+\varepsilon}$. Define $X = \sum_{j=1}^n X_j$ and $f(\mathbf{X}) = iX + \frac{1}{n} e^{-iX}$. We obviously have $\mathbb{E} X = 0$ and $\mathbb{E} X^2 = n^{2\varepsilon}$. For $c = \pm 1$, we have

$$\begin{aligned} \mathbb{E} e^{icX} &= (\mathbb{E} e^{icX_1})^n = \left(\frac{1}{2} e^{-in^{-1/2+\varepsilon}} + \frac{1}{2} e^{in^{-1/2+\varepsilon}} \right)^n \\ &= \left(1 - \frac{1}{2} n^{-1+2\varepsilon} + O(n^{-3/2+3\varepsilon}) \right)^n = e^{-n^{2\varepsilon}/2 + O(n^{-1/2+3\varepsilon})}. \end{aligned}$$

Using $e^{f(\mathbf{X})} = e^{iX} + \frac{1}{n} + O(\frac{1}{n^2})$ we have $\mathbb{E} e^{f(\mathbf{X})} = \frac{1}{n} + O(\frac{1}{n^2})$. Now let us apply Theorem 3.3. We have $\mathbb{E} f(\mathbf{X}) = \frac{1}{n} e^{-n^{2\varepsilon}/2 + o(1)}$ and $\mathbb{V} f(\mathbf{X}) = -n^{2\varepsilon} + o(1)$. Therefore $e^{\mathbb{E} f + \frac{1}{2} \mathbb{V} f} = e^{-\frac{1}{2} n^{2\varepsilon} + o(1)}$. In the error term of (3.4) we have $\alpha_k = O(n^{-1/2+\varepsilon})$ and $\Delta_{jk} = O(n^{-2+2\varepsilon})$. So $e^{\mathbb{E} f + \frac{1}{2} \mathbb{V} f}$ is very much smaller than $\mathbb{E} e^{f(\mathbf{X})}$ even though $L' = o(1)$. In a later paper we will investigate a wide class of complex functions for which a theorem similar to Theorem 3.3 is true without the factor $e^{\frac{1}{2} \text{Var } \Im f(\mathbf{X})}$.

3.2 Smooth and transformed functions

In the case of smooth functions, the parameters α_j and Δ_{jk} can be bounded in terms of derivatives or other measures such as Lipchitz constants. For our applications in Section 4, it will suffice to have continuous differentiability.

For $\rho > 0$, define

$$U_n(\rho) = \{\mathbf{x} \in \mathbb{R}^n \mid |x_j| \leq \rho \text{ for } 1 \leq j \leq n\}.$$

If f is a function one of whose arguments is x , then f_x is the partial derivative $\partial f / \partial x$, and similarly for notations like f_{xy} . If the arguments are a subscripted list, like x_1, \dots, x_n , we will further abbreviate f_{x_j} to f_j and $f_{x_j x_k}$ to f_{jk} . The notations $\|\cdot\|_1$, $\|\cdot\|_2$ and $\|\cdot\|_\infty$ have their usual meanings as vector norms and the corresponding induced matrix norms. For a matrix $A = (a_{jk})$ we will also use $\|A\|_{\max} = \max_{jk} |a_{jk}|$ but note that it is not submultiplicative.

Lemma 3.6.

(a) Let L be the closed line segment $[x_1, x_2] \subseteq \mathbb{R}$ and let S be its interior minus a countable set of points. Suppose that the function $f : L \rightarrow \mathbb{C}$ is continuous, and that f_x exists and is bounded in S . Then

$$|f(x_2) - f(x_1)| \leq |x_2 - x_1| \sup_{x \in S} |f_x(x)|.$$

(b) Let R be the closed rectangle $[x_1, x_2] \times [y_1, y_2] \subseteq \mathbb{R}^2$ and let S be its interior minus a countable set of lines. Suppose that the function $f : R \rightarrow \mathbb{C}$ is continuous and f_x exists and is continuous. Moreover assume that f_{xy} exists and is bounded in S . Then

$$|f(x_1, y_1) - f(x_1, y_2) - f(x_2, y_1) + f(x_2, y_2)| \leq |x_2 - x_1| |y_2 - y_1| \sup_{(x,y) \in S} |f_{xy}(x, y)|.$$

Proof. The conditions we have given are sufficient to imply that

$$f(x_2) - f(x_1) = \int_L^{(\text{HK})} f_x(x) dx$$

in case (a) and

$$f(x_1, y_1) - f(x_1, y_2) - f(x_2, y_1) + f(x_2, y_2) = \int_{[y_1, y_2]}^{(\text{HK})} \left(\int_{[x_1, x_2]}^{(\text{HK})} f_{xy}(x, y) dx \right) dy,$$

in case (b), where we have used the Henstock–Kurzweil (gauge) integral [2, Thm. 4.7]. The claims now follow readily. \square

Note that in part (a) we did not require that f_x is continuous, and in part (b) we did not require that f_y or f_{xy} are continuous. The lemma is not true if “countable” is replaced by “measure zero”. In the following we will adopt more stringent conditions on derivatives than Lemma 3.6 allows, leaving the generalizations to the reader.

Corollary 3.7. *Let $B = [a_1, b_1] \times \cdots \times [a_n, b_n] \subseteq \mathbb{R}^n$. Suppose that $f : B \rightarrow \mathbb{C}$ is continuous. Then, provided the suprema exist,*

(a) *If f is continuously differentiable in the interior $\text{int } B$ of B , then, for $1 \leq j \leq n$,*

$$\alpha_j(f, B) \leq (b_j - a_j) \sup_{\mathbf{x} \in \text{int } B} |f_j(\mathbf{x})|.$$

(b) *If f is twice continuously differentiable in $\text{int } B$, then, for $1 \leq j < k \leq n$,*

$$\Delta_{jk}(f, B) \leq (b_j - a_j)(b_k - a_k) \sup_{\mathbf{x} \in \text{int } B} |f_{jk}(\mathbf{x})|.$$

Proof. This follows immediately from Lemma 3.6, noting that line segments or rectangles in the boundary of B are limits of line segments or rectangles not in the boundary. \square

In the case of a transformed cuboid, it is convenient to be able to bound $\|\boldsymbol{\alpha}\|_\infty$, $\boldsymbol{\alpha}^T \Delta \boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^T \Delta^2 \boldsymbol{\alpha}$ in terms of the derivatives in the original coordinates. We will only treat the case of uniformly bounded derivatives.

If $B \subseteq \mathbb{R}^n$ is some set and $f : B \rightarrow \mathbb{C}$ is twice differentiable in some open set containing B , define the matrix $H(f, B) = (h_{jk})$, where, provided the suprema exist, $h_{jk} = \sup_{\mathbf{y} \in B} |f_{jk}(\mathbf{y})|$.

Lemma 3.8. *For some $\rho > 0$, let $B = U_n(\rho)$. Suppose that $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a differentiable transformation and let J_T denote its Jacobian matrix. Let $S \subseteq \mathbb{R}^m$ be an open set that contains $T(\text{int } B)$. Suppose $f : T(B) \cup S \rightarrow \mathbb{C}$ is continuous, and define $\tilde{f} : B \rightarrow \mathbb{C}$ by $\tilde{f}(\mathbf{x}) = f(T(\mathbf{x}))$. Write $\boldsymbol{\alpha} = \boldsymbol{\alpha}(\tilde{f}, B)$ and $\Delta = (\Delta_{jk}) = \Delta(\tilde{f}, B)$. Then*

(a) Suppose that f is continuously differentiable in S with $|f_j(\mathbf{y})| \leq m_1$ for $\mathbf{y} \in T(\text{int } B)$ and $1 \leq j \leq n$. Then

$$\|\boldsymbol{\alpha}\|_\infty \leq 2\rho m_1 \sup_{\mathbf{x} \in \text{int } B} \|J_T(\mathbf{x})\|_1.$$

(b) Suppose that f is twice continuously differentiable in S with $\|H(f, T(\text{int } B))\|_\infty \leq m_2$. Then

$$\begin{aligned} \boldsymbol{\alpha}^T \Delta \boldsymbol{\alpha} &\leq 4\rho^2 n m_2 \|\boldsymbol{\alpha}\|_\infty^2 \sup_{\mathbf{x} \in \text{int } B} (\|J_T(\mathbf{x})\|_1 \|J_T(\mathbf{x})\|_\infty) \quad \text{and} \\ \boldsymbol{\alpha}^T \Delta^2 \boldsymbol{\alpha} &\leq 16\rho^4 n m_2^2 \|\boldsymbol{\alpha}\|_\infty^2 \sup_{\mathbf{x} \in \text{int } B} (\|J_T(\mathbf{x})\|_1 \|J_T(\mathbf{x})\|_\infty)^2. \end{aligned}$$

Proof. Suppose $J_T = (t_{jk}(\mathbf{x}))$. Observe that for $\mathbf{x} \in \text{int } B$

$$\begin{aligned} \tilde{f}_j(\mathbf{x}) &= \sum_{r=1}^m t_{rj}(\mathbf{x}) f_r(T(\mathbf{x})), \\ \tilde{f}_{jk}(\mathbf{x}) &= \sum_{r=1}^m \sum_{s=1}^n t_{rj}(\mathbf{x}) f_{rs}(T(\mathbf{x})) t_{sk}(\mathbf{x}). \end{aligned}$$

From Corollary 3.7 for function \tilde{f} , we get

$$\alpha_j \leq 2\rho m_1 \sup_{\mathbf{x} \in \text{int } B} \sum_{r=1}^m |t_{rj}(\mathbf{x})|,$$

which is equivalent to part (a), and

$$\Delta_{jk} \leq 4\rho^2 \sup_{\mathbf{x} \in \text{int } B} \sum_{r=1}^m \sum_{s=1}^n |t_{rj}(\mathbf{x})| h_{rs} |t_{sk}(\mathbf{x})|,$$

where $H(f, T(\text{int } B)) = (h_{rs})$. Note that the expression $\sum_{r=1}^m \sum_{s=1}^n |t_{rj}(\mathbf{x})| h_{rs} |t_{sk}(\mathbf{x})|$ of the right hand side is the (j, k) element of $(\hat{J}_T)^T H(f, T(\text{int } B)) \hat{J}_T$, where $\hat{J}_T = (|t_{rs}|)$. Claim (b) now follows on recalling that the ∞ -norm is submultiplicative. \square

4 Truncated gaussian measures

In this section explore the application of Theorem 3.3 to the case where the distribution of \mathbf{X} is a truncated gaussian. This is the case that has occurred in the most applications so far.

It will often be convenient to approximate the expectation and pseudovariance of a complex function of a truncated gaussian to their values for the unrestricted gaussian. The following gives a general principle.

Lemma 4.1. *Let A be an $n \times n$ symmetric positive-definite real matrix. Let $f : \mathbb{R}^n \rightarrow \mathbb{C}$ be a measurable function satisfying*

$$|f(\mathbf{x})| \leq e^{\frac{b}{n} \mathbf{x}^T A \mathbf{x}} \quad (4.1)$$

for all $\mathbf{x} \in \mathbb{R}^n$ and some $b \geq 0$. Let $\mathbf{X} : \mathbb{R} \rightarrow \mathbb{R}$ be a random variable with density

$$\pi^{-n/2} |A|^{1/2} e^{-\mathbf{x}^T A \mathbf{x}}.$$

Suppose Ω is a measurable subset of \mathbb{R}^n and define $p = \text{Prob}(\mathbf{X} \notin \Omega)$. Then, if $p \leq \frac{3}{4}$ and $n \geq b + b^2$, we have

$$|\mathbb{E}(f(\mathbf{X}) \mid \mathbf{X} \in \Omega) - \mathbb{E} f(\mathbf{X})| \leq 15 e^{b/2} p^{1-b/n}.$$

Moreover, for $p \leq \frac{3}{4}$ and $n \geq 2b + 4b^2$, we have

$$|\mathbb{V}(f(\mathbf{X}) \mid \mathbf{X} \in \Omega) - \mathbb{V} f(\mathbf{X})| \leq 112 e^b p^{1-2b/n}.$$

Proof. By linear transform we can assume that $A = \frac{1}{2}I$ and that $|f(\mathbf{x})| \leq e^{\frac{b}{2n} \mathbf{x}^T \mathbf{x}}$. Let μ denote the measure with density $\pi^{-n/2} e^{-\frac{1}{2} \mathbf{x}^T \mathbf{x}}$, which is the gaussian measure defined by \mathbf{X} after transformation. From the definition of expectation,

$$\mathbb{E}(f(\mathbf{X}) \mid \mathbf{X} \in \Omega) - \mathbb{E} f(\mathbf{X}) = (1-p)^{-1} \left(p \int_{\mathbb{R}^n} f(\mathbf{x}) d\mu - \int_{\mathbb{R}^n - \Omega} f(\mathbf{x}) d\mu \right).$$

For any $r > 0$, since $\text{Prob}(\mathbf{X} \notin \Omega \wedge |\mathbf{X}| \leq r) \leq p$, we can bound

$$\int_{\mathbb{R}^n - \Omega} |f(\mathbf{x})| d\mu \leq p \sup_{|\mathbf{x}| \leq r} |f(\mathbf{x})| + \int_{|\mathbf{x}| \geq r} |f(\mathbf{x})| d\mu.$$

Consequently we have

$$|\mathbb{E}(f(\mathbf{X}) \mid \mathbf{X} \in \Omega) - \mathbb{E} f(\mathbf{X})| \leq (1-p)^{-1} \left(p e^{\frac{b}{2n} r^2} + \int_{|\mathbf{x}| \geq r} e^{\frac{b}{2n} \mathbf{x}^T \mathbf{x}} d\mu + p \int_{\mathbb{R}^n} e^{\frac{b}{2n} \mathbf{x}^T \mathbf{x}} d\mu \right).$$

The second integral is easily calculated to be $(1 - b/n)^{-n/2}$, provided $n > b$. The first integral has no closed form; it is $(1 - b/n)^{-n/2} F_n((1 - b/n)r^2)$, where $F_n(u)$ denotes the upper tail of the χ^2 -distribution with n degrees of freedom. From [28, (4.3)] we have that $F_n(n + 2u^{1/2}n^{1/2} + 2u) \leq e^{-u}$ for any $u \geq 0$. Consequently, if we put $r^2 = (n + 2u^{1/2}n^{1/2} + 2u)/(1 - b/n)$, we find for any $u \geq 0$ and $n > b$ that

$$\begin{aligned} & |\mathbb{E}(f(\mathbf{X}) \mid \mathbf{X} \in \Omega) - \mathbb{E} f(\mathbf{X})| \\ & \leq (1-p)^{-1} (p e^{b(1/2 + (u/n)^{1/2} + u/n)/(1-b/n)} + (1 - b/n)^{-n/2} (p + e^{-u})). \end{aligned}$$

To obtain the version in the theorem statement, use

$$u = (1 - b/n) \ln(1/p) + \frac{b \sqrt{4(n - b) \ln(1/p) - b^2}}{2n},$$

which satisfies the equation $b(1/2 + (u/n)^{1/2} + u/n)/(1 - b/n) = -u + \ln(1/p)$. The conditions $p \leq \frac{3}{4}$ and $n \geq b + b^2$ imply that the argument of the square root is positive. Now note that for $n \geq b + b^2, b \geq 0$ the function $(1 - b/n)^{-n/2} e^{-b/2} < e^{1/4}$ is increasing in b and nonincreasing in n , so $(1 - b/n)^{-n/2} e^{-b/2} < e^{1/4}$. Applying this bound and also $u \geq (1 - b/n) \ln(1/p)$ completes the proof of the first part.

For the second part, we have

$$\begin{aligned} \mathbb{V}(f(\mathbf{X}) \mid \mathbf{X} \in \Omega) - \mathbb{V}f(\mathbf{X}) &= \mathbb{E}(f(\mathbf{X})^2 \mid \mathbf{X} \in \Omega) - \mathbb{E}f(\mathbf{X})^2 \\ &\quad - (\mathbb{E}(f(\mathbf{X}) \mid \mathbf{X} \in \Omega) - \mathbb{E}f(\mathbf{X})) (\mathbb{E}f(\mathbf{X}) + \mathbb{E}(f(\mathbf{X}) \mid \mathbf{X} \in \Omega)). \end{aligned}$$

Note from above that $|\mathbb{E}f(\mathbf{X})| \leq \mathbb{E}|f(\mathbf{X})| \leq (1 - b/n)^{-n/2}$. Using the definition of p , we have $|\mathbb{E}(f(\mathbf{X}) \mid \mathbf{X} \in \Omega)| \leq (1 - p)^{-1} \int_{\Omega} |f(\mathbf{X})| d\mu \leq 4(1 - b/n)^{-n/2}$. Now apply the first part of the lemma to $f(\mathbf{X})$ and $f(\mathbf{X})^2$, as well as the bound $(1 - b/n)^{-n/2} e^{-b/2} < e^{1/4}$ used earlier. This completes the proof. \square

Lemma 4.1 is not useful for exponential functions on account of condition (4.1). However since (4.1) is satisfied for all polynomials, the lemma becomes useful in conjunction with Theorem 3.3 for estimating $\mathbb{E}e^f$ when f has polynomial growth. For convenience, we give the theorem of Isserlis [25] (see [18] for a treatment in modern notation) that tells us how to compute the expectations of polynomials with respect to a multivariate normal distribution.

Theorem 4.2. *Let A be a positive-definite real symmetric matrix of order n and let $\mathbf{X} = (X_1, \dots, X_n)$ be a random variable with the normal density $\pi^{-n/2} |A|^{-1/2} e^{-\mathbf{x}^T A \mathbf{x}}$. Let $\Sigma = (\sigma_{jk}) = (2A)^{-1}$ be the corresponding covariance matrix. Consider a product $Z = X_{j_1} X_{j_2} \cdots X_{j_k}$, where the subscripts do not need to be distinct. If k is odd, then $\mathbb{E}Z = 0$. If k is even, then*

$$\mathbb{E}Z = \sum_{(i_1, i_2), (i_2, i_3), \dots, (i_{k-1}, i_k)} \sigma_{j_{i_1} j_{i_2}} \cdots \sigma_{j_{i_{k-1}} j_{i_k}},$$

where the sum is over all unordered partitions of $\{1, 2, \dots, k\}$ into $k/2$ disjoint unordered pairs. The number of terms in the sum is $(k-1)(k-3) \cdots 3 \cdot 1$.

The following are examples of Theorem 4.2.

$$\begin{aligned} \mathbb{E}X_1^2 &= \sigma_{11} & \mathbb{E}X_1^4 &= 3\sigma_{11}^2 \\ \mathbb{E}X_1^2 X_2^2 &= \sigma_{11}\sigma_{22} + 2\sigma_{12}^2 & \mathbb{E}X_1^2 X_2 X_3 &= \sigma_{11}\sigma_{23} + 2\sigma_{12}\sigma_{13} \\ \mathbb{E}X_1 X_2 X_3 X_4 &= \sigma_{12}\sigma_{34} + \sigma_{13}\sigma_{24} + \sigma_{14}\sigma_{23} & \mathbb{E}X_1^6 &= 15\sigma_{11}^3 \end{aligned}$$

4.1 Truncated gaussian measures of full rank

Theorem 4.3. *Let $c_1, c_2, c_3, \varepsilon, \rho_1, \rho_2, \phi_1, \phi_2$ be nonnegative real constants with $c_1, \varepsilon > 0$. Let A be an $n \times n$ positive-definite symmetric real matrix and let T be a real matrix such that $T^T A T = I$. Let Ω be a measurable set such that $U_n(\rho_1) \subseteq T^{-1}(\Omega) \subseteq U_n(\rho_2)$, and let $f : \mathbb{R}^n \rightarrow \mathbb{C}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}$ and $h : \Omega \rightarrow \mathbb{C}$ be measurable functions. We make the following assumptions.*

- (a) $c_1(\log n)^{1/2+\varepsilon} \leq \rho_1 \leq \rho_2$.
- (b) For $\mathbf{x} \in T(U_n(\rho_1))$, $2\rho_1 \|T\|_1 |f_j(\mathbf{x})| \leq \phi_1 n^{-1/2}$ for each j .
- (c) For $\mathbf{x} \in \Omega$, $\Re f(\mathbf{x}) \leq g(\mathbf{x})$. For $\mathbf{x} \in T(U_n(\rho_2))$, $2\rho_2 \|T\|_1 |g_j(\mathbf{x})| \leq \phi_2 n^{-1/2}$ for each j .
- (d) $|f(\mathbf{x})|, |g(\mathbf{x})| \leq n^{c_3} e^{c_2 \mathbf{x}^T A \mathbf{x}/n}$ for $\mathbf{x} \in \mathbb{R}^n$.

Let \mathbf{X} be a random variable with the normal density $\pi^{-n/2} |A|^{1/2} e^{-\mathbf{x}^T A \mathbf{x}}$. Then, provided $\mathbb{E} f(\mathbf{X})$ and $\mathbb{E} g(\mathbf{X})$ are finite and h is bounded in Ω ,

$$\int_{\Omega} e^{-\mathbf{x}^T A \mathbf{x} + f(\mathbf{x}) + h(\mathbf{x})} d\mathbf{x} = (1 + K) \pi^{n/2} |A|^{-1/2} e^{\mathbb{E} f(\mathbf{X})},$$

where, for some constant C depending only on $c_1, c_2, c_3, \varepsilon$,

$$|K| \leq C \left(e^{\frac{1}{8}\phi_1^2 + e^{-\rho_1^2/2}} - 1 + (2e^{\frac{1}{8}\phi_2^2 + e^{-\rho_1^2/2}} - 2 + \sup_{\mathbf{x} \in \Omega} |e^{h(\mathbf{x})} - 1|) e^{\mathbb{E}(g(\mathbf{X}) - \Re f(\mathbf{X}))} \right).$$

In particular, if $n \geq (1 + c_2)^2$ and $\rho_1^2 \geq 7 + 2c_2 + (3 + 4c_3) \log n$, we can take $C = 1$.

Proof. We will use Lemma 6.2 repeatedly to combine error terms. Change variables by $\mathbf{x} = T\mathbf{y}$. Since $|T| = |A|^{-1/2}$, we have

$$\int_{\Omega} e^{-\mathbf{x}^T A \mathbf{x} + f(\mathbf{x}) + h(\mathbf{x})} d\mathbf{x} = |A|^{-1/2} \int_{T^{-1}(\Omega)} e^{-\mathbf{y}^T \mathbf{y} + f(T\mathbf{y}) + h(T\mathbf{y})} d\mathbf{y}.$$

Suppose $\rho \geq c_1(\log n)^{1/2+\varepsilon}$ and let $F : U_n(\rho) \rightarrow \mathbb{C}$ be measurable and such that $|F(\mathbf{x})| \leq n^{c_3} e^{c_2 \mathbf{x}^T \mathbf{x}/n}$ for $\mathbf{x} \in \mathbb{R}^n$ and $\|\alpha(F, U_n(\rho))\|_{\infty} \leq \phi n^{-1/2}$. By Theorem 3.3(a),

$$\int_{U_n(\rho)} e^{-\mathbf{y}^T \mathbf{y} + F(\mathbf{y})} d\mathbf{y} = (1 + K') e^{\mathbb{E}(F(\mathbf{Y}) | \mathbf{Y} \in U_n(\rho))} \int_{U_n(\rho)} e^{-\mathbf{y}^T \mathbf{y}} d\mathbf{y},$$

where $|K'| \leq e^{\frac{1}{8}\phi^2} - 1$ and \mathbf{Y} has the normal density $\pi^{-n/2} e^{-\mathbf{y}^T \mathbf{y}}$.

Define $p = \text{Prob}(\mathbf{Y} \notin U_n(\rho))$. By standard bounds on the tail of the normal distribution, $p \leq n e^{-\rho^2} / (1 + \rho)$. Under our assumptions, there is $n_0 = n_0(c_1, c_2, c_3, \varepsilon)$ such that for

$n \geq n_0$, we have $p \leq \frac{3}{4}$, $n \geq c_2 + c_2^2$ and $15n^{c_3}e^{c_2/2}p^{1-c_2/n} \leq e^{-\rho^2/2}$. Those three conditions are enough that we can apply Lemma 4.1 to the function $n^{-c_3}F(\mathbf{y})$ to conclude that

$$\int_{U_n(\rho)} e^{-\mathbf{y}^T \mathbf{y} + F(\mathbf{y})} d\mathbf{y} = (1 + K'') \pi^{n/2} e^{\mathbb{E}F(\mathbf{Y})}, \quad (4.2)$$

where $|K''| \leq e^{\frac{1}{8}\phi^2 + e^{-\rho^2/2}} - 1$.

We can finish the proof by applying (4.2) to each of the functions $f(T\mathbf{y})$ and $g(T\mathbf{y})$. For $n < n_0$ we can increase C to make the theorem hold, so assume $n \geq n_0$. By Lemma 3.8 we have $\|\alpha(f(T\mathbf{y}), U_n(\rho_1))\|_\infty \leq n^{-1/2}\phi_1$ and $\|\alpha(g(T\mathbf{y}), U_n(\rho_2))\|_\infty \leq n^{-1/2}\phi_2$. Now we have

$$\begin{aligned} & \int_{\Omega} e^{f(T\mathbf{y}) + h(T\mathbf{y}) - \mathbf{y}^T \mathbf{y}} d\mathbf{y} \\ &= \int_{U_n(\rho_1)} e^{f(T\mathbf{y}) - \mathbf{y}^T \mathbf{y}} d\mathbf{y} + A \left(\int_{U_n(\rho_2)} - \int_{U_n(\rho_1)} \right) e^{g(T\mathbf{y}) - \mathbf{y}^T \mathbf{y}} d\mathbf{y} \\ & \quad + A' \sup_{\mathbf{x} \in \Omega} |e^{h(\mathbf{x})} - 1| \int_{\mathbb{R}^n} e^{g(T\mathbf{y}) - \mathbf{y}^T \mathbf{y}} d\mathbf{y} \quad \text{for } |A|, |A'| \leq 1 \\ &= \pi^{n/2} e^{\mathbb{E}f(\mathbf{X})} (1 + K_1) + K_2 \pi^{n/2} e^{\mathbb{E}g(\mathbf{X})} + K_3 \pi^{n/2} e^{\mathbb{E}g(\mathbf{X})}, \end{aligned} \quad (4.3)$$

where we have $|K_1| \leq e^{\frac{1}{8}\phi_1^2 + e^{-\rho_1^2/2}} - 1$, $|K_2| \leq 2e^{\frac{1}{8}\phi_2^2 + e^{-\rho_1^2/2}} - 2$ and $|K_3| \leq \sup_{\mathbf{x} \in \Omega} |e^{h(\mathbf{x})} - 1|$. Finally note that $|e^{\mathbb{E}f(\mathbf{X})}| = e^{\mathbb{E}\Re f(\mathbf{X})}$; the theorem follows.

To establish the final claim, it will suffice to show that for $\rho^2 \geq 7 + 2c_2 + (3 + 4c_3) \log n$ we can prove (4.2) with $n_0 = (1 + c_2)^2$. Obviously $n \geq (1 + c_2)^2$ implies that $n \geq c_2 + c_2^2$, and it also implies that $1 - c_2/n \geq \frac{3}{4}$. The bounds $\rho^2 \geq 7 + 3 \log n$ and $p \leq ne^{-\rho^2}/(1 + \rho)$ imply that $p \leq \frac{3}{4}$ and also that $p \leq e^{-\rho^2}n/(1 + \sqrt{7})$. Combining these bounds produces the third required inequality $15n^{c_3}e^{c_2/2}p^{1-c_2/n} \leq e^{-\rho^2/2}$, completing the proof. \square

Theorem 4.4. *Let $c_1, c_2, c_3, \varepsilon, \rho_1, \rho_2, \phi_1, \phi_2$ be nonnegative real constants with $c_1, \varepsilon > 0$. Let A be an $n \times n$ positive-definite symmetric real matrix and let T be a real matrix such that $T^T A T = I$. Let Ω be a measurable set such that $U_n(\rho_1) \subseteq T^{-1}(\Omega) \subseteq U_n(\rho_2)$, and let $f : \mathbb{R}^n \rightarrow \mathbb{C}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}$ and $h : \Omega \rightarrow \mathbb{C}$ be measurable functions. We make the following assumptions.*

- (a) $c_1(\log n)^{1/2+\varepsilon} \leq \rho_1 \leq \rho_2$.
- (b) For $\mathbf{x} \in T(U_n(\rho_1))$, $2\rho_1 \|T\|_1 |f_j(\mathbf{x})| \leq \phi_1 n^{-1/3} \leq \frac{2}{3}$ for $1 \leq j \leq n$ and $4\rho_1^2 \|T\|_1 \|T\|_\infty \|H(f, T(U_n(\rho_1)))\|_\infty \leq \phi_1 n^{-1/3}$.
- (c) For $\mathbf{x} \in \Omega$, $\Re f(\mathbf{x}) \leq g(\mathbf{x})$. For $\mathbf{x} \in T(U_n(\rho_2))$, either
 - (i) $2\rho_2 \|T\|_1 |g_j(\mathbf{x})| \leq (2\phi_2)^{3/2} n^{-1/2}$ for $1 \leq j \leq n$, or

$$(ii) \ 2\rho_2 \|T\|_1 |g_j(\mathbf{x})| \leq \phi_2 n^{-1/3} \text{ for } 1 \leq j \leq n \text{ and} \\ 4\rho_2^2 \|T\|_1 \|T\|_\infty \|H(g, T(U_n(\rho_2)))\|_\infty \leq \phi_2 n^{-1/3}.$$

$$(d) \ |f(\mathbf{x})|, |g(\mathbf{x})| \leq n^{c_3} e^{c_2 \mathbf{x}^T A \mathbf{x}/n} \text{ for } \mathbf{x} \in \mathbb{R}^n.$$

Let \mathbf{X} be a random variable with the normal density $\pi^{-n/2} |A|^{1/2} e^{-\mathbf{x}^T A \mathbf{x}}$. Then, provided $\mathbb{V}f(\mathbf{X})$ and $\mathbb{V}g(\mathbf{X})$ are finite and h is bounded in Ω ,

$$\int_{\Omega} e^{-\mathbf{x}^T A \mathbf{x} + f(\mathbf{x}) + h(\mathbf{x})} d\mathbf{x} = (1 + K) \pi^{n/2} |A|^{-1/2} e^{\mathbb{E}f(\mathbf{X}) + \frac{1}{2} \mathbb{V}f(\mathbf{X})},$$

where, for some constant C depending only on $c_1, c_2, c_3, \varepsilon$,

$$|K| \leq C e^{\frac{1}{2} \text{Var} \Im f(\mathbf{X})} \left(e^{\phi_1^3 + e^{-\rho_1^2/2}} - 1 \right. \\ \left. + (2e^{\phi_2^3 + e^{-\rho_1^2/2}} - 2 + \sup_{\mathbf{x} \in \Omega} |e^{h(\mathbf{x})} - 1|) e^{\mathbb{E}(g(\mathbf{X}) - \Re f(\mathbf{X})) + \frac{1}{2} (\text{Var} g(\mathbf{X}) - \text{Var} \Re f(\mathbf{X}))} \right).$$

In particular, if $n \geq (1 + 2c_2)^2$ and $\rho_1^2 \geq 15 + 4c_2 + (3 + 8c_3) \log n$, we can take $C = 1$.

Proof. We will divide the integral in the same fashion as (4.3), and will use estimate (4.2) again. We also need a similar estimate using Theorem 3.3(b). Lemma 6.2 will be used to combine error terms.

Suppose $\rho \geq c_1(\log n)^{1/2+\varepsilon}$ and let $F : U_n(\rho) \rightarrow \mathbb{C}$ be measurable and such that $|F(\mathbf{x})| \leq n^{c_3} e^{c_2 \mathbf{x}^T \mathbf{x}}$ for $\mathbf{x} \in \mathbb{R}^n$, and for $\mathbf{x} \in T(U_n(\rho))$, $\|\alpha(F, U_n(\rho))\|_\infty \leq \phi n^{-1/3} \leq \frac{2}{3}$ and $\|\Delta(F, U_n(\rho))\|_\infty \leq \phi n^{-1/3}$. By Theorem 3.3(b),

$$\int_{U_n(\rho)} e^{-\mathbf{y}^T \mathbf{y} + F(\mathbf{y})} d\mathbf{y} = (1 + K' e^{\frac{1}{2} \text{Var}(\Im F(\mathbf{Y}) | \mathbf{Y} \in U_n(\rho))}) \\ \times e^{\mathbb{E}(F(\mathbf{Y}) | \mathbf{Y} \in U_n(\rho)) + \frac{1}{2} \mathbb{V}(F(\mathbf{Y}) | \mathbf{Y} \in U_n(\rho))} \int_{U_n(\rho)} e^{-\mathbf{y}^T \mathbf{y}} d\mathbf{y},$$

where $|K'| \leq e^{\phi^3} - 1$ and \mathbf{Y} has the normal density $\pi^{-n/2} e^{-\mathbf{y}^T \mathbf{y}}$. Similarly to the proof of Theorem 4.3, we can apply Lemma 4.1 to conclude that there is a constant $n_0 = n_0(c_1, c_2, c_3, \varepsilon)$ such that for $n \geq n_0$,

$$\int_{U_n(\rho)} e^{-\mathbf{y}^T \mathbf{y} + F(\mathbf{y})} d\mathbf{y} = (1 + K'' e^{\frac{1}{2} \text{Var} \Im F(\mathbf{Y})}) \pi^{n/2} e^{\mathbb{E}F(\mathbf{Y}) + \frac{1}{2} \mathbb{V}F(\mathbf{Y})}, \quad (4.4)$$

where $|K''| \leq e^{\phi^3 + e^{-\rho^2/2}} - 1$.

Now consider the expansion given by (4.3). By condition (b) and Lemma 3.8, we have $\|\alpha(f(T\mathbf{y}), U_n(\rho_1))\|_\infty \leq n^{-1/3} \rho_1 \leq \frac{2}{3}$, and $\|\Delta(f(T\mathbf{y}), U_n(\rho_1))\|_\infty \leq n^{-1/3} \rho_1$. Consequently, by (4.4), we have for $n \geq n_0$ that

$$\int_{U_n(\rho_1)} e^{-\mathbf{y}^T \mathbf{y} + f(T\mathbf{y})} d\mathbf{y} = (1 + K_1 e^{\frac{1}{2} \text{Var} \Im f(\mathbf{Y})}) \pi^{n/2} e^{\mathbb{E}f(\mathbf{Y}) + \frac{1}{2} \mathbb{V}f(\mathbf{Y})}, \quad (4.5)$$

where $|K_1| \leq e^{\phi_1^3 + e^{-\rho_1^2/2}} - 1$.

For the second part of (4.3), we need separate consideration of the two cases of condition (c). In case (ii) we can apply (4.2) to $g(T\mathbf{y})$ to obtain

$$\left(\int_{U_n(\rho_2)} - \int_{U_n(\rho_1)} \right) e^{g(T\mathbf{y}) - \mathbf{y}^T \mathbf{y}} d\mathbf{y} = K'_2 \pi^{n/2} e^{\mathbb{E}g(\mathbf{X})}, \quad (4.6)$$

where $|K'_2| \leq 2(e^{\phi_2^3 + e^{-\rho_1^2/2}} - 1)$ for $n \geq n_0$. In case (ii) we can assume $\phi_2 n^{-1/3} \leq \frac{2}{3}$ or else case (i) applies. Then (4.4) gives for $n \geq n_0$ that

$$\left(\int_{U_n(\rho_2)} - \int_{U_n(\rho_1)} \right) e^{g(T\mathbf{y}) - \mathbf{y}^T \mathbf{y}} d\mathbf{y} = K''_2 \pi^{n/2} e^{\mathbb{E}g(\mathbf{X}) + \frac{1}{2} \mathbb{V}g(\mathbf{X})}, \quad (4.7)$$

where $|K''_2| \leq 2(e^{\phi_2^3 + e^{-\rho_1^2/2}} - 1)$. Since $e^{\frac{1}{2} \mathbb{V}g(\mathbf{X})} \geq 1$ (g being real), we can write both (4.6) and (4.7) as $K_2 \pi^{n/2} e^{\mathbb{E}g(\mathbf{X}) + \frac{1}{2} \mathbb{V}g(\mathbf{X})}$, where $|K_2| \leq \min(|K'_2|, |K''_2|) \leq 2(e^{\phi_2^3 + e^{-\rho_1^2/2}} - 1)$.

The third part of (4.3) is bounded in modulus by $\sup_{\mathbf{x} \in \Omega} |e^{h(\mathbf{x})} - 1| \pi^{n/2} e^{\mathbb{E}g(\mathbf{X})}$ just as in Theorem 4.3. Adding the three parts, and noting that C can be increased to cover the finite number of cases when $n < n_0$, the theorem follows.

The final claim is proved essentially as in the previous theorem. \square

Remark 4.5. Note that $U_n(\rho_1) \subseteq T^{-1}(\Omega) \subseteq U_n(\rho_2)$ is implied by $U_n(\rho_1 \| T \|_\infty) \subseteq \Omega \subseteq U_n(\rho_2 \| T^{-1} \|_\infty^{-1})$, so the latter condition could be used instead of the former.

4.2 Truncated gaussian measures of less than full rank

Many enumeration problems have generating functions with symmetries that lead to singular quadratic forms. As an example, which we will work in more detail in Section 4.3, regular tournaments are counted by the constant term of $\prod_{1 \leq j < k \leq n} (x_j/x_k + x_k/x_j)$, which is invariant under multiplication of each variable by the same constant [32]. Expanding at the saddle-point gives the quadratic form $\sum_{1 \leq j < k \leq n} (\theta_j - \theta_k)^2$, which is invariant in the direction $(1, 1, \dots, 1)$. By conditioning on the value of one variable, or the sum of the variables, we can restrict the integral to a subspace of codimension 1. In other problems the codimension can be higher. Here we provide a general technique that expands such integrals to full dimension, so that the techniques of the previous subsection can be applied.

If $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a linear operator, let $\ker T = \{\mathbf{x} \in \mathbb{R}^n \mid T\mathbf{x} = 0\}$. If L is a linear subspace of \mathbb{R}^n , let L^\perp be the orthogonal complement.

Lemma 4.6. Let $Q, W : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be linear operators such that $\ker Q \cap \ker W = \{\mathbf{0}\}$ and $\text{span}(\ker Q, \ker W) = \mathbb{R}^n$. Let n^\perp denote the dimension of $\ker Q$. Suppose $\Omega \subseteq \mathbb{R}^n$ and $F : \Omega \cap Q(\mathbb{R}^n) \rightarrow \mathbb{C}$. For any $\rho > 0$, define

$$\Omega_\rho = \{\mathbf{x} \in \mathbb{R}^n \mid Q\mathbf{x} \in \Omega \text{ and } W\mathbf{x} \in U_n(\rho)\}.$$

Then, if the integrals exist,

$$\int_{\Omega \cap Q(\mathbb{R}^n)} F(\mathbf{y}) d\mathbf{y} = (1 - K)^{-1} \pi^{-n^\perp/2} |Q^T Q + W^T W|^{1/2} \int_{\Omega_\rho} F(Q\mathbf{x}) e^{-\mathbf{x}^T W^T W \mathbf{x}} d\mathbf{x},$$

where

$$0 \leq K < \min(1, ne^{-\rho^2/\kappa^2}), \quad \kappa = \sup_{W\mathbf{x} \neq 0} \frac{\|W\mathbf{x}\|_\infty}{\|W\mathbf{x}\|_2} \leq 1.$$

Moreover, if $U_n(\rho_1) \subseteq \Omega \subseteq U_n(\rho_2)$ for some $\rho_2 \geq \rho_1 > 0$ then

$$U_n\left(\min\left(\frac{\rho_1}{\|Q\|_\infty}, \frac{\rho}{\|W\|_\infty}\right)\right) \subseteq \Omega_\rho \subseteq U_n\left(\|P\|_\infty \rho_2 + \|R\|_\infty \rho\right)$$

for any linear operators $P, R : \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that $PQ + RW$ is equal to the identity operator on \mathbb{R}^n .

Proof. The bounds on Ω_ρ follow directly from the definition of Ω_ρ : for the lower bound, we use $\|Q\mathbf{x}\|_\infty \leq \|Q\|_\infty \|\mathbf{x}\|_\infty \leq \rho_1$ and $\|W\mathbf{x}\|_\infty \leq \|W\|_\infty \|\mathbf{x}\|_\infty \leq \rho$; for the upper bound, apply $\|\mathbf{x}\|_\infty \leq \|P\|_\infty \|Q\mathbf{x}\|_\infty + \|R\|_\infty \|W\mathbf{x}\|_\infty$.

Due to assumptions on $\ker Q$ and $\ker W$, we can find some invertible linear operator $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that $T(\ker Q) = (T(\ker W))^\perp$. Substituting $\mathbf{x} = T^{-1}\hat{\mathbf{x}}$, we get that

$$\int_{\Omega_\rho} F(Q\mathbf{x}) e^{-\mathbf{x}^T W^T W \mathbf{x}} d\mathbf{x} = |T|^{-1} \int_{T(\Omega_\rho)} F(\hat{Q}\hat{\mathbf{x}}) e^{-\hat{\mathbf{x}}^T \hat{W}^T \hat{W} \hat{\mathbf{x}}} d\hat{\mathbf{x}},$$

where $\hat{Q} = QT^{-1}$, $\hat{W} = WT^{-1}$. Note that $\ker \hat{Q} = T(\ker Q)$ and $\ker \hat{W} = T(\ker W)$. Consider an orthonormal basis consisting of $n - n_\perp$ vectors that span $\ker \hat{W}$ and n_\perp vectors that span $\ker \hat{Q}$. Thus $\mathbf{x}' \in \mathbb{R}^n$ is represented as $(\mathbf{x}_Q, \mathbf{x}_W) \in \ker \hat{W} \oplus \ker \hat{Q}$. The quadratic form with matrix $\hat{Q}^T \hat{Q} + \hat{W}^T \hat{W}$ acts separately on the orthogonal subspaces $\ker \hat{W}$ and $\ker \hat{Q}$, therefore $|\hat{Q}^T \hat{Q} + \hat{W}^T \hat{W}|^{1/2} = |T|^{-1} |Q^T Q + W^T W|^{1/2}$ is equal to the product of the Jacobian determinants of the linear maps corresponding to the restrictions of \hat{Q} to $\ker \hat{W}$ and \hat{W} to $\ker \hat{Q}$. Then we have

$$\begin{aligned} \int_{T(\Omega_\rho)} F(\hat{Q}\hat{\mathbf{x}}) e^{-\hat{\mathbf{x}}^T \hat{W}^T \hat{W} \hat{\mathbf{x}}} d\hat{\mathbf{x}} &= \int_{\ker \hat{W} \cap T(\Omega_\rho)} F(\hat{Q}\mathbf{x}_Q) d\mathbf{x}_Q \times \int_{\ker \hat{Q} \cap T(\Omega_\rho)} e^{-\mathbf{x}_W^T \hat{W}^T \hat{W} \mathbf{x}_W} d\mathbf{x}_W \\ &= |Q^T Q + W^T W|^{-1/2} |T| \int_{\Omega \cap Q(\mathbb{R}^n)} F(\mathbf{y}) d\mathbf{y} \times \int_{U_n(\rho) \cap W(\mathbb{R}^n)} e^{-\mathbf{z}^T \mathbf{z}} d\mathbf{z}, \end{aligned}$$

where the last integral would be $\pi^{n_{\perp}/2}$ except for the restricted domain. Thus the claim follows with $K = \text{Prob}(\mathbf{X} \notin U_n(\rho))$, where \mathbf{X} is a normal variable on $W(\mathbb{R}^n)$ with density $\pi^{-n_{\perp}/2} e^{-\mathbf{z}^T \mathbf{z}}$. The cube $U_n(\rho)$ intersects $W(\mathbb{R}^n)$ in a convex polytope whose facets are intersections of $W(\mathbb{R}^n)$ with the facets of $U_n(\rho)$. By the definition of κ the perpendicular distance from the origin to a facet of $U_n(\rho) \cap W(\mathbb{R}^n)$ is at least equal to $\inf_{\text{facet of } U_n(\rho) \cap W(\mathbb{R}^n)} \|\mathbf{z}\|_{\infty} / \kappa = \rho / \kappa$. Since there are at most $2n$ such facets, we have that $K \leq 2n\pi^{-1/2} \int_{\rho/\kappa}^{\infty} e^{-x^2} dx \leq ne^{-\rho^2/\kappa^2}$. \square

Using unbounded regions, we get the following corollary.

Corollary 4.7. *Let $\Omega = \mathbb{R}^n$ and assumptions of Lemma 4.6 hold. For any linear subspace $L \subseteq \mathbb{R}^n$ such that $L \cap \ker Q = \{0\}$ and $\text{span}(L, \ker Q) = \mathbb{R}^n$, define a random variable \mathbf{Y}_L taking values in L with density proportional to $e^{-\mathbf{y}^T Q^T Q \mathbf{y}}$. Then $\mathbb{E} F(Q\mathbf{Y}_L)$ does not depend on the choice of L and is equal to $\mathbb{E} F(Q\mathbf{X}_W)$, where \mathbf{X}_W is the random variable taking values in \mathbb{R}^n with density proportional to $e^{-\mathbf{x}^T (Q^T Q + W^T W) \mathbf{x}}$.*

Proof. Observe that

$$\mathbb{E} F(Q\mathbf{Y}_L) = \frac{\int_{Q(\mathbb{R}^n)} F(\mathbf{z}) e^{-\mathbf{z}^T \mathbf{z}} d\mathbf{z}}{\int_{Q(\mathbb{R}^n)} e^{-\mathbf{z}^T \mathbf{z}} d\mathbf{z}}.$$

To complete the proof, we use Lemma 4.6 with $\rho \rightarrow \infty$, which implies $K \rightarrow 0$. \square

4.3 Example: regular tournaments

The enumeration of regular tournaments make a good example for the demonstration of how Lemma 4.6 can be used to reduce an integral to a form for which Theorem 4.3 applies. We recall that a regular tournament is a complete digraph in which the in-degree is equal to the out-degree at each vertex. Let $RT(n)$ be the number of labelled regular tournaments with n vertices. It is clear that $RT(n) = 0$ if n is even. The following formula was given for the first time in [32]:

Theorem 4.8. *For odd $n \rightarrow \infty$*

$$RT(n) = \left(1 + O(n^{-1/2+\varepsilon})\right) \left(\frac{2^{n+1}}{\pi n}\right)^{(n-1)/2} n^{1/2} e^{-1/2} \quad (4.8)$$

for any $\varepsilon > 0$.

Proof. Observe that $RT(n)$ is equal to the constant term of the generating function $\prod_{1 \leq j < k \leq n} (x_j/x_k + x_k/x_j)$. Using contours $x_j = e^{i\theta_j}$, we get by Cauchy theorem that

$$RT(n) = \frac{2^{n(n-1)/2}}{(2\pi)^n} \text{Int}, \quad \text{Int} = \int_{U_n(\pi)} \prod_{1 \leq j < k \leq n} \cos(\theta_j - \theta_k) d\boldsymbol{\theta}.$$

The next step, which we omit here and refer to [32, Sect. 3], is to show that for odd $n \rightarrow \infty$

$$\text{Int} = (1 + O(e^{-cn^{2\varepsilon}})) 2^n \pi \int_{U_{n-1}(n^{-1/2+\varepsilon})} \prod_{1 \leq j < k \leq n} \cos(\theta_j - \theta_k) d\boldsymbol{\theta}'$$

for some $c > 0$, where the integration is with respect to $\boldsymbol{\theta}' = (\theta_1, \dots, \theta_{n-1})$ with $\theta_n = 0$.

Now let $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^n$ and

$$\begin{aligned} \Omega &= U_n(n^{-1/2+\varepsilon}), \quad F(\mathbf{x}) = \prod_{1 \leq j < k \leq n} \cos(x_j - x_k), \\ Q\mathbf{x} &= \mathbf{x} - x_n \mathbf{1}, \quad W\mathbf{x} = \frac{1}{\sqrt{2n}}(x_1 + \dots + x_n) \mathbf{1}, \\ P\mathbf{x} &= \mathbf{x} - \frac{1}{n}(x_1 + \dots + x_n) \mathbf{1}, \quad R\mathbf{x} = \sqrt{\frac{2}{n}} \mathbf{x}. \end{aligned}$$

We observe $F(Q\mathbf{x}) = F(\mathbf{x})$ and apply Lemma 4.6 with $\rho = \frac{1}{\sqrt{2}}n^\varepsilon$ to get

$$\begin{aligned} \int_{U_{n-1}(n^{-1/2+\varepsilon})} \prod_{1 \leq j < k \leq n} \cos(\theta_j - \theta_k) d\boldsymbol{\theta}' &= \int_{\Omega \cap Q(\mathbb{R}^n)} F(\mathbf{y}) d\mathbf{y} \\ &= (1 + O(e^{-c'n^{2\varepsilon}})) \pi^{-1/2} 2^{-1/2} n \int_{\Omega_\rho} F(\mathbf{x}) e^{-\frac{1}{2}(x_1 + \dots + x_n)^2} d\mathbf{x} \end{aligned}$$

for some $c' > 0$. We obtain also that $U_n(\frac{1}{2}n^{-1/2+\varepsilon}) \subseteq \Omega_\rho \subseteq U_n(3n^{-1/2+\varepsilon})$. By Taylor's theorem, we can expand

$$F(\mathbf{x}) e^{-\frac{1}{2}(x_1 + \dots + x_n)^2} = \exp\left(-\frac{n}{2} \mathbf{x}^T \mathbf{x} - \frac{1}{12} \sum_{1 \leq j < k \leq n} (x_j - x_k)^4 + O(n^{-1+6\varepsilon})\right).$$

Define \mathbf{X} to be the gaussian random variable with density $(2\pi)^{-n/2} n^{n/2} e^{-\frac{n}{2} \mathbf{x}^T \mathbf{x}}$ and let $f(\mathbf{x}) = -\frac{1}{12} \sum_{1 \leq j < k \leq n} (x_j - x_k)^4$. Then $\mathbb{E} f(\mathbf{X}) = -\frac{(n-1)}{2n}$ and $\partial f / \partial x_j = O(n^{-3/2+4\varepsilon})$ for $\mathbf{x} \in \Omega_\rho$ and $1 \leq j \leq n$. Now apply Theorem 4.3 with $A = \frac{n}{2}I$, $T = \sqrt{\frac{2}{n}}I$, $\rho_1, \rho_2 = O(n^\varepsilon)$, $\phi_1, \phi_2 = O(n^{-1/2+4\varepsilon})$ and $g(\mathbf{x}) = f(\mathbf{x})$ to find that

$$\int_{\Omega_\rho} F(\mathbf{x}) e^{-\frac{1}{2}(x_1 + \dots + x_n)^2} d\mathbf{x} = (1 + O(n^{-1+8\varepsilon})) 2^{n/2} \pi^{n/2} n^{-n/2} e^{-1/2}.$$

Formula (4.8) follows. □

Although we used an old theorem here for illustrative purposes, it is worth nothing that the same method can be used to enumerate tournaments according to score sequence over a very wide range of scores, well beyond that achieved in [14]. The details will appear separately.

An example of how Lemma 4.6 can be applied in conjunction with Theorem 4.4 is the enumeration of bipartite graphs, which we will cover in Section 5.

4.4 The case of weakly dependent components

In order to apply Theorems 4.3 and 4.4 to particular examples, we need to know that there exists some linear transformation T such that $T^T A T = I$ and which satisfies good bounds on $\|T\|_1$, $\|T\|_\infty$ (and $\|T^{-1}\|_\infty$). In this subsection we give a general recipe for finding T in the case when diagonal elements of A are of the same order while off-diagonal elements are relatively small. Equivalently, the components of the corresponding gaussian random variable are weakly dependent.

As was mentioned in Sections 4.2 and 4.3 sometimes we have that A is a positive-semidefinite matrix with non-trivial kernel and the region of integration lies in some linear subspace of \mathbb{R}^n . Then, using Lemma 4.6, we can reduce it to the integration over a region of full dimension with modified quadratic form $A + W^T W$ which is non-singular. For such purposes we also need analogous estimates for a linear transform T satisfying $T^T(A + W^T W)T = I$. There is a large flexibility of choosing W in general. One strategy is to make $A + W^T W$ close to some diagonal matrix and proceed as in the case of full dimension. Alternatively, when A is close to some diagonal matrix D but entries of $W^T W$ are always big in comparison with entries of $A - D$ it turned out to be better to choose W in one particular way as described below.

If D is a positive-semidefinite matrix, we denote by $D^{1/2}$ the positive-semidefinite square root and, in the case of nonsingularity, by $D^{-1/2}$ the positive-definite inverse square root. Let

$$A_D = A + D^{1/2} P_D D^{1/2},$$

where P_D is the linear operator that projects orthogonally onto $D^{1/2}(\ker A)$. Assuming that D is not singular, note that $A_D \mathbf{x} = A \mathbf{x}_\parallel + D \mathbf{x}_\perp$, where

$$\mathbf{x}_\perp = D^{-1/2} P_D D^{1/2} \mathbf{x} \in \ker A, \quad \mathbf{x}_\parallel = \mathbf{x} - \mathbf{x}_\perp = D^{-1/2} (I - P_D) D^{1/2} \mathbf{x}.$$

In the case of $\ker A = \{\mathbf{0}\}$ we have $A_D = A$ and $\mathbf{x}_\parallel = \mathbf{x}$.

Lemma 4.9. *Let D be an $n \times n$ real diagonal matrix with $d_{\min} = \min_j d_{jj} > 0$ and $d_{\max} = \max_j d_{jj}$. Recall the norm $\|\cdot\|_{\max}$ defined in Section 3.2. Let A be a real symmetric positive-semidefinite $n \times n$ matrix with*

$$\|A - D\|_{\max} \leq \frac{rd_{\min}}{n} \quad \text{and} \quad \mathbf{x}^T A \mathbf{x} \geq \gamma \mathbf{x}_{\parallel}^T D \mathbf{x}_{\parallel} = \gamma \mathbf{x}^T D^{1/2} (I - P_D) D^{1/2} \mathbf{x}$$

for some $1 \geq \gamma > 0$, $r > 0$ and all $\mathbf{x} \in \mathbb{R}^n$. Let n_{\perp} denote the dimension of $\ker A$. Then the following are true.

$$(a) \quad \|A_D - A\|_{\infty} \leq rn_{\perp} d_{\max}^{1/2} d_{\min}^{1/2}, \quad \|A_D - A\|_{\max} \leq \frac{r^2 n_{\perp} d_{\min}}{n} \quad \text{and} \quad n_{\perp} \leq r^2.$$

(b) A_D is symmetric and positive-definite. Moreover,

$$\|A_D^{-1}\|_{\infty} \leq \frac{r + \gamma}{\gamma d_{\min}} \quad \text{and} \quad \|A_D^{-1} - D^{-1}\|_{\max} \leq \frac{(r + \gamma)r}{\gamma n d_{\min}} (1 + rn_{\perp}).$$

(d) There exists a matrix T such that $T^T A_D T = I$ and

$$\|T\|_1, \|T\|_{\infty} \leq \frac{r + \gamma^{1/2}}{\gamma^{1/2} d_{\min}^{1/2}} \quad \text{and} \quad \|T^{-1}\|_1, \|T^{-1}\|_{\infty} \leq \left(\frac{(r + 1)(r + \gamma^{1/2})}{\gamma^{1/2}} + rn_{\perp} \right) d_{\max}^{1/2}.$$

Proof. Let \mathbf{y} be a unit vector of $D^{1/2}(\ker A)$. Then $D^{1/2}\mathbf{y} = (D - A)D^{-1/2}\mathbf{y}$, so by assumption and using $\|\mathbf{y}\|_1 \leq n^{1/2}\|\mathbf{y}\|_2 = n^{1/2}$, we find that

$$\|D^{1/2}\mathbf{y}\|_{\infty} \leq \frac{rd_{\min}}{n} \|D^{-1/2}\mathbf{y}\|_1 \leq \frac{rd_{\min}^{1/2}}{n^{1/2}}, \quad \|D^{1/2}\mathbf{y}\|_1 \leq d_{\max}^{1/2} \|\mathbf{y}\|_1 \leq n^{1/2} d_{\max}^{1/2}.$$

Consequently we get $\|D^{1/2}\mathbf{y}\mathbf{y}^T D^{1/2}\|_{\infty} \leq rd_{\min}^{1/2} d_{\max}^{1/2}$ and $\|D^{1/2}\mathbf{y}\mathbf{y}^T D^{1/2}\|_{\max} \leq \frac{r^2 d_{\min}}{n}$. If $\{\mathbf{y}_1, \dots, \mathbf{y}_{n_{\perp}}\}$ is a full set of orthonormal vectors of $D^{1/2}(\ker A)$ then $P_D = \sum_{j=1}^{n_{\perp}} \mathbf{y}_j \mathbf{y}_j^T$ and $A_D - A = \sum_{j=1}^{n_{\perp}} D^{1/2} \mathbf{y}_j \mathbf{y}_j^T D^{1/2}$ which implies first two estimates of part (a). The last estimate of part (a) follows from the observation that the trace of $(A - D)^2$ is at most $r^2 d_{\min}^2$ and at least $n_{\perp} d_{\min}^2$.

For any $\mathbf{x} \in \mathbb{R}^n$ and $t \in \mathbb{R}$ we have

$$\begin{aligned} \|(t^2 D + A_D)\mathbf{x}\|_{\infty} &= \|(t^2 + 1)D\mathbf{x} + (A - D)\mathbf{x}\|_{\infty} \geq (t^2 + 1)d_{\min} \|\mathbf{x}\|_{\infty} - \frac{rd_{\min}}{n} \|\mathbf{x}\|_1 \\ &\geq (t^2 + 1)d_{\min} \|\mathbf{x}\|_{\infty} - \frac{rd_{\min}}{n^{1/2}} \|\mathbf{x}\|_2. \end{aligned}$$

Note that eigenvalues of $D^{-1/2} \hat{A} D^{-1/2} = D^{-1/2} A D^{-1/2} + P_D$ are positive eigenvalues of $D^{-1/2} A D^{-1/2}$ that are at least γ , plus n_{\perp} extra eigenvalues equal 1. Putting $\mathbf{y} = D^{1/2}\mathbf{x}$, we get that

$$\begin{aligned} \|(t^2 D + A_D)\mathbf{x}\|_{\infty} &\geq \frac{d_{\min}^{1/2}}{n^{1/2}} \|(t^2 I + D^{-1/2} A D^{-1/2})\mathbf{y}\|_2 \geq \frac{(t^2 + \gamma)d_{\min}^{1/2}}{n^{1/2}} \|\mathbf{y}\|_2 \\ &\geq \frac{(t^2 + \gamma)d_{\min}^{1/2}}{n^{1/2}} \|P_D \mathbf{y}\|_2 \geq \frac{(t^2 + \gamma)d_{\min}}{n^{1/2}} \|\mathbf{x}\|_2. \end{aligned}$$

Adding $t^2 + \gamma$ times the first inequality to r times the second, we find that

$$(t^2 + \gamma + r)\|(t^2 D + A_D)\mathbf{x}\|_\infty \geq (t^2 + \gamma)(t^2 + 1)d_{\min}\|\mathbf{x}\|_\infty, \quad (4.9)$$

which implies for $t = 0$ the first estimate of part (b). The second estimate follows from

$$\|A_D^{-1} - D^{-1}\|_{\max} = \|A_D^{-1}(A_D - D)D^{-1}\|_{\max} \leq d_{\min}^{-1}\|A_D^{-1}\|_\infty \|A_D - D\|_{\max}.$$

Let $B = D^{-1/2}AD^{-1/2}$. Note that B satisfies conditions of Lemma 4.9 with I playing role of D . Namely, $\|B - I\|_{\max} \leq \frac{1}{d_{\min}}\|A - D\|_{\max} \leq r/n$ and $\mathbf{x}^T B \mathbf{x} \geq \gamma \mathbf{x}^T (I - P_{\ker B}) \mathbf{x}$, where $P_{\ker B}$ is the orthogonal projector onto $\ker B$. From [16, p. 133] we have that

$$(B + P_{\ker B})^{-1/2} = \frac{2}{\pi} \int_0^\infty (t^2 I + B + P_{\ker B})^{-1} dt.$$

Using the bound (4.9) with B playing role of A , we find that

$$\|(t^2 I + B + P_{\ker B})^{-1}\|_\infty \leq \frac{t^2 + \gamma + r}{(t^2 + \gamma)(t^2 + 1)}.$$

Performing the integral gives

$$\|(B + P_{\ker B})^{-1/2}\|_\infty \leq \frac{r + \gamma^{1/2} + \gamma}{\gamma^{1/2} + \gamma} \leq \frac{r + \gamma^{1/2}}{\gamma^{1/2}}.$$

From the eigensystems we see that $(B + P_{\ker B})^k$ acts the same as B^k on vectors in $\ker B^\perp$ and preserves vectors in $\ker B$. Consequently,

$$(B + P_{\ker B})^{1/2} = B(B + P_{\ker B})^{-1/2} + P_{\ker B}.$$

Claim (a) with B playing role of A gives $\|P_{\ker B}\|_\infty \leq rn_\perp$. Using also $\|B\|_\infty \leq r + 1$, we get that

$$\|(B + P_{\ker B})^{1/2}\|_\infty \leq \frac{(r+1)(r+\gamma^{1/2})}{\gamma^{1/2}} + rn_\perp.$$

In order to prove claim (c) we can take $T = D^{-1/2}(B + P_{\ker B})^{-1/2}$, which indeed satisfies $T^T A_D T = I$. For $p \in \{1, \infty\}$,

$$\|T\|_p = \|D^{-1/2}(B + P_{\ker B})^{-1/2}\|_p \leq d_{\min}^{-1/2} \|(B + P_{\ker B})^{-1/2}\|_p,$$

and similarly

$$\|T^{-1}\|_p \leq d_{\max}^{1/2} \|(B + P_{\ker B})^{1/2}\|_p.$$

Recalling that $(B + P_{\ker B})^{-1/2}$ and $(B + P_{\ker B})^{1/2}$ are symmetric, claim (c) follows. \square

Remark 4.10. Actually, Lemma 4.9 is valid not only when D is a diagonal matrix but also when D is any symmetric positive-definite $n \times n$ matrix, with $\|D^{1/2}\|_\infty^2$ playing the role of d_{\max} and $\|D^{-1/2}\|_\infty^{-2}$ playing the role of d_{\min} . The proof of this generalization is identical.

5 Graphs with given degrees

In this section we will demonstrate the use of our theory to obtain new results on graphs with given degrees. We will generalise the problem as follows.

Let $H = (H^+, H^-)$ be a pair of fixed (simple) edge-disjoint graphs on vertices $V = \{1, \dots, n\}$. We will not notationally distinguish graphs from their edge-sets. Let $N_H(\mathbf{d})$ be the number of graphs on V which have vertex degrees $\mathbf{d} = (d_1, \dots, d_n)$, include H^+ as a subgraph, and are edge-disjoint from H^- . The generating function for N_H is

$$\begin{aligned} F_{\mathbf{d}, H}(x_1, \dots, x_n) &= \sum_{d_1, \dots, d_n} N_H(\mathbf{d}) x_1^{d_1} \cdots x_n^{d_n} \\ &= \prod_{\{j, k\} \in H^+} x_j x_k \prod_{\{j, k\} \notin H^+ \cup H^-} (1 + x_j x_k). \end{aligned} \quad (5.1)$$

From this it follows that

$$N_H(\mathbf{d}) = \frac{1}{(2\pi i)^n} \oint \cdots \oint \frac{F_{\mathbf{d}, H}(z_1, \dots, z_n)}{z_1^{d_1+1} \cdots z_n^{d_n+1}} dz_1 \cdots dz_n, \quad (5.2)$$

where each contour circles the origin once anticlockwise.

The value of $N_{(\emptyset, \emptyset)}(\mathbf{d})$ was estimated by McKay and Wormald [36] when d_1, \dots, d_n are large (approximately a constant fraction of n) and not very far from equal. McKay [31] later extended this to the case of nonempty H , provided $H^+ \cup H^-$ has at most $n^{1+\varepsilon}$ edges and maximum degree at most $n^{1/2+\varepsilon}$. Meanwhile, Barvinok and Hartigan [5] extended the case of $H = (\emptyset, \emptyset)$ to a much wider range of degrees.

Our definition also includes the bipartite case. Let $V_1 = \{1, \dots, n_1\}$, $V_2 = \{n_1+1, \dots, n_1+n_2\}$ be a partition of V into two disjoint subsets. Define $\tilde{E} = \binom{V_1}{2} \cup \binom{V_2}{2}$; that is, the complement of the complete bipartite graph with parts V_1 and V_2 . If $\tilde{E} \subseteq H^-$, then $N_H(\mathbf{d})$ is a count of bipartite graphs.

Canfield and McKay [8] estimated $\tilde{N}_{(\emptyset, \emptyset)}(\mathbf{d})$ in the semiregular case, which was later extended to more irregular degree sequences by Barvinok and Hartigan [5]. The case where $H^+ \cup H^- \neq \emptyset$ was treated by Greenhill and McKay [15] if \mathbf{d} is not far from semiregular.

We will generalise all these results. Changing variables in (5.2) with $z_j = e^{\beta_j + i\theta_j}$, and defining

$$p_{jk} = \begin{cases} 0 & \text{if } j = k \text{ or } \{j, k\} \in H^-; \\ 1 & \text{if } \{j, k\} \in H^+; \\ \frac{e^{\beta_j + \beta_k}}{1 + e^{\beta_j + \beta_k}} & \text{otherwise,} \end{cases} \quad (5.3)$$

we have

$$N_H(\mathbf{d}) = C_{\mathbf{d},H} \int_{U_n(\pi)} G_{\mathbf{d},H}(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (5.4)$$

where

$$\begin{aligned} U_n(t) &= \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid -t \leq x_j \leq t \text{ for } 1 \leq j \leq n\}, \\ C_{\mathbf{d},H} &= \frac{\prod_{\{j,k\} \in H^+} e^{\beta_j + \beta_k} \prod_{\{j,k\} \notin H^+ \cup H^-} (1 + e^{\beta_j + \beta_k})}{(2\pi)^n e^{d_1\beta_1 + \dots + d_n\beta_n}}, \\ G_{\mathbf{d},H}(\boldsymbol{\theta}) &= \frac{\prod_{\{j,k\} \in H^+} e^{i(\theta_j + \theta_k)} \prod_{\{j,k\} \notin H^+ \cup H^-} (1 + p_{jk}(e^{i(\theta_j + \theta_k)} - 1))}{e^{i(d_1\theta_1 + \dots + d_n\theta_n)}}. \end{aligned} \quad (5.5)$$

Equation (5.4) is valid for any radii $\{e^{\beta_j}\}$, but in order to estimate the integral we need its value to be concentrated in a small region where the integrand is not too oscillatory. There are also symmetries to take into account. The most obvious is that $G_{\mathbf{d},H}(\theta_1, \dots, \theta_n) = G_{\mathbf{d},H}(\theta_1 + \pi, \dots, \theta_n + \pi)$. In the bipartite case we also have that $G_{\mathbf{d},H}(\theta_1, \dots, \theta_n) = G_{\mathbf{d},H}(\theta_1 + t, \dots, \theta_{n_1} + t, \theta_{n_1+1} - t, \dots, \theta_n - t)$ for any t . Other symmetries can occur if the complement of $H^+ \cup H^-$ is disconnected, but we will not consider those cases here.

A good choice of radii is that which makes the contours pass together through the saddle point on the positive real axis. This gives the equations

$$\sum_{k=1}^n p_{jk} = d_j, \quad (1 \leq j \leq n), \quad (5.6)$$

in which case we have

$$C_{\mathbf{d},H} = (2\pi)^{-n} \prod_{1 \leq j < k \leq n} p_{jk}^{-p_{jk}} (1 - p_{jk})^{-(1-p_{jk})},$$

where $0^0 = 1$ as usual. There is no comprehensive theory about when $\{\beta_j\}$ exist to satisfy (5.6), but much is known in the cases $H = (\emptyset, \emptyset)$ and $H = (\emptyset, \tilde{E})$, which will suffice for us here.

In the case $H = (\emptyset, \emptyset)$, a unique solution for $\{\beta_j\}$ exists if \mathbf{d} lies in the interior of the polytope defined by the Erdős-Gallai inequalities [5, 13, 39]. The corresponding values $\{p_{jk}\}$ have an important property: if we generate a random graph, where for each j, k , there is an edge from vertex j to vertex k with probability p_{jk} , such choices made independently, then the probability of any graph depends only on its degree sequence and, moreover, the expected degree sequence is \mathbf{d} . Conversely, the equal-probability condition implies that the edge probabilities are related as in (5.3) and the expected degree condition

implies that (5.6) holds [13]. Following [12], we call this the β -model of random graph corresponding to \mathbf{d} .

For the basic bipartite case $H = (\emptyset, \tilde{E})$, for any solution β_1, \dots, β_n and any $b \in \mathbb{R}$, $\beta_1 - b, \dots, \beta_{n_1} - b, \beta_{n_1+1} + b, \dots, \beta_n + b$ is also a solution, but note that the resulting values of $\{p_{jk}\}$ remain the same. With this caveat, the solution exists and is unique if \mathbf{d} lies in the relative interior of the polytope of bipartite degree sequences [40]. Similarly to before, if we generate a random bipartite graph with parts V_1, V_2 and edges chosen independently with probabilities p_{jk} , then the probability of every bipartite graph with parts V_1, V_2 depends only on its degree sequence and the expected degree sequence is \mathbf{d} . We will call this the *bipartite β -model* and note that it is also called the *Rasch model* [40].

In the following subsections we will determine asymptotic values for $N_H(\mathbf{d})$ and $\tilde{N}_H(\mathbf{d})$ using the same range of degree sequences as allowed by Barvinok and Hartigan [5], but with nontrivial H . This will enable us to prove that the distribution of edges within a constant or slowly-increasing set of vertex pairs is asymptotically equal to that for the corresponding β -model. This strengthens the result of Chatterjee et al. [12] that graphs with given degrees converge in the sense of graph limits to the graphon defined by the β -model, under some simple conditions.

In Section 5.3 we show that the number of edges within an arbitrary set of vertex pairs is concentrated near the same value for random graphs with given degrees and random graphs in the corresponding β -model. This considerably strengthens similar results of Barvinok and Hartigan [3, 5]

In all cases, we will not present the best results our theory allows so as to keep this example focussed. We will say more about that at the end of Section 5.2.

5.1 General graphs

Throughout this subsection, we will define λ_{jk} to be the value of p_{jk} in the solution of (5.3) subject to (5.6) in the case $H = (\emptyset, \emptyset)$.

We now follow Barvinok and Hartigan [5] by requiring that \mathbf{d} is δ -tame for some $\delta > 0$, which means that $\delta \leq \lambda_{jk} \leq 1 - \delta$ for all $j \neq k$. Chatterjee et al. [12] showed that δ -tameness follows if \mathbf{d} is not too close to the boundary of the Erdős-Gallai polytope. Barvinok and Hartigan provide a useful sufficient condition.

Lemma 5.1 ([5]). *Let $0 < \alpha < \beta < 1$ satisfy $(\alpha + \beta)^2 < 4\alpha$. Then if $\alpha(n - 1) < d_j < \beta(n - 1)$ for $1 \leq j \leq n$ and n is large enough, there is some $\delta > 0$ such that \mathbf{d} is δ -tame.*

Define the $n \times n$ symmetric matrix A by

$$\boldsymbol{\theta}^T A \boldsymbol{\theta} = \frac{1}{2} \sum_{j < k} \lambda_{jk} (1 - \lambda_{jk}) (\theta_j + \theta_k)^2.$$

For each j , let s_j be the number of times vertex j occurs in $H^+ \cup H^-$ and define $s_{\max} = \max_{j=1}^n s_j$, $S = \frac{1}{2} \sum_{j=1}^n s_j$ and $S_2 = \sum_{j=1}^n s_j^2$. Also define the following function, which arises from Taylor expansion of $G_{\mathbf{d}, H}(\boldsymbol{\theta}) + \boldsymbol{\theta}^T A \boldsymbol{\theta}$ about the origin.

$$\begin{aligned} f_H(\boldsymbol{\theta}) = & i \sum_{\{j,k\} \in H^+} (1 - \lambda_{jk}) (\theta_j + \theta_k) - i \sum_{\{j,k\} \in H^-} \lambda_{jk} (\theta_j + \theta_k) \\ & + \frac{1}{2} \sum_{\{j,k\} \in H^+ \cup H^-} \lambda_{jk} (1 - \lambda_{jk}) (\theta_j + \theta_k)^2 \\ & - \frac{1}{6} i \sum_{\{j,k\} \notin H^+ \cup H^-} \lambda_{jk} (1 - \lambda_{jk}) (1 - 2\lambda_{jk}) (\theta_j + \theta_k)^3 \\ & + \frac{1}{24} \sum_{\{j,k\} \notin H^+ \cup H^-} \lambda_{jk} (1 - \lambda_{jk}) (1 - 6\lambda_{jk} + 6\lambda_{jk}^2) (\theta_j + \theta_k)^4. \end{aligned} \quad (5.7)$$

Now we can state our main enumeration result, and the resulting estimate of

$$P_H(\mathbf{d}) = \frac{N_H(\mathbf{d})}{N_{(\emptyset, \emptyset)}(\mathbf{d})},$$

which is the probability that a uniform random graph with degrees \mathbf{d} contains H^+ and is disjoint from H^- .

Theorem 5.2. *Let \mathbf{d} be δ -tame for some $\delta > 0$. Define $\{r_j\}, \{\lambda_{jk}\}, A, s_{\max}, S_2, f_H$ as above, and suppose that $s_{\max} \leq c_1 n^{1/6}$ and $S_2 \leq c_2 n$ for constants c_1, c_2 . Let \mathbf{X} be a random variable with the normal density $\pi^{-n/2} |A|^{1/2} e^{-\mathbf{x}^T A \mathbf{x}}$. Then, for any $\varepsilon > 0$, there is a constant $c = c(\delta, \varepsilon, c_1, c_2)$ such that*

$$N_H(\mathbf{d}) = 2 \pi^{n/2} C_{\mathbf{d}, H} |A|^{-1/2} e^{\mathbb{E} \Re f_H(\mathbf{X}) - \frac{1}{2} \mathbb{E} (\Im f_H(\mathbf{X}))^2} (1 + K), \quad (5.8)$$

where $|K| \leq e^{c(1+s_{\max}^3)n^{-1/2+\varepsilon}} - 1$. Moreover,

$$P_H(\mathbf{d}) = (1 + K') \prod_{\{j,k\} \in H^+} \lambda_{jk} \prod_{\{j,k\} \in H^-} (1 - \lambda_{jk})$$

where $|K'| \leq e^{cS_2/n + c(1+s_{\max}^3)n^{-1/2+\varepsilon}} - 1$.

Note that formula (5.8) in the case of $H = (\emptyset, \emptyset)$ matches [5, Thm. 1.4] apart from the error term. The formula for $P_H(\mathbf{d})$, absent the error term, is the same as for the β -model. The formula for $P_H(\mathbf{d})$ is given more precisely in [31], but only for the near-regular degree

sequences considered there. It considerably strengthens [12, Thm. 1], at least for δ -tame degree sequences.

Proof. For the duration of the proof, the implied constant in each $O(\cdot)$ expression depends only on $\delta, \varepsilon, c_1, c_2$. We begin with a sequence of lemmas. Define $\Omega = U_n(\log n/n^{1/2})$.

Lemma 5.3. *For any $k > 0$,*

$$\int_{U_n(\pi)} G_{\mathbf{d},H}(\boldsymbol{\theta}) d\boldsymbol{\theta} = 2 \int_{\Omega} G_{\mathbf{d},H}(\boldsymbol{\theta}) d\boldsymbol{\theta} + O(n^{-k}) \int_{\Omega} |G_{\mathbf{d},(\emptyset,\emptyset)}(\boldsymbol{\theta})| d\boldsymbol{\theta}$$

Proof. This is proved by the same method used in [5, Thm. 8.1], with only a small change in their Lemma 8.4 to allow for the $o(n)$ factors for each j , of the form $|1 + \lambda_{jk}(e^{i(\theta_j + \theta_k)} - 1)|$, that appear in $|G_{\mathbf{d},(\emptyset,\emptyset)}|$ but not in $|G_{\mathbf{d},H}|$. \square

Lemma 5.4. *Let D be the diagonal matrix with the same diagonal as A . Then for some constant a_1 we have $\|A^{-1} - D^{-1}\|_{\max} \leq a_1 n^{-2}$. Furthermore, there exists a matrix T with $T^T A T = I$ and some constants a_2, a_3 such that $\|T\|_1, \|T\|_{\infty} \leq a_2 n^{-1/2}$ and $\|T^{-1}\|_{\infty} \leq a_3 n^{1/2}$.*

Proof. From the definition of A we have $\|A - D\|_{\max} \leq \frac{1}{8}$. Also, for any \mathbf{x} we have

$$\mathbf{x}^T A \mathbf{x} \geq \frac{1}{2} \delta (1 - \delta) \sum_{j < k} (x_j + x_k)^2 \geq \frac{1}{2} \delta (1 - \delta) (n - 2) \mathbf{x}^T \mathbf{x},$$

where we used the fact that the least eigenvalue of the matrix of the quadratic form $\sum_{j < k} (x_j + x_k)^2$ is $n - 2$. Taking into account that

$$\max D_{jj} \leq \frac{1}{2} \delta (1 - \delta) (n - 1) \leq \frac{1}{8} (n - 1) \quad \text{and} \quad \min D_{jj} \geq \frac{1}{2} \delta (1 - \delta) (n - 1),$$

we apply Lemma 4.9 with $r = n/(4\delta(1 - \delta)(n - 1))$ and $\gamma = 4\delta(1 - \delta)\frac{n-2}{n-1}$ to complete the proof. \square

Lemma 5.5. *We have*

$$\begin{aligned} \mathbb{E} f_H(\mathbf{X}) &= \mathbb{E} f_{(\emptyset,\emptyset)}(\mathbf{X}) + O(S/n) = O(1), \\ \text{Var } \Re f_H(\mathbf{X}) &= \text{Var } \Re f_{(\emptyset,\emptyset)}(\mathbf{X}) + O(S_2/n^2) = O(1/n), \quad \text{and} \\ \text{Var } \Im f_H(\mathbf{X}) &= \text{Var } \Im f_{(\emptyset,\emptyset)}(\mathbf{X}) + O(S_2/n) = O(1). \end{aligned}$$

Proof. Consider the covariance matrix $(2A)^{-1} = (\sigma_{jk})$ and the random variable $\mathbf{X} = (X_1, \dots, X_n)$ defined in the theorem. By Lemma 5.4, we have $\sigma_{jj} = O(n^{-1})$ and $\sigma_{jk} =$

$O(n^{-2})$ for all $j \neq k$. Lemma 4.2 now tells us that any odd monomial in X_1, \dots, X_n has mean 0, and that for $p, q \in \mathbb{N}$ and $j \neq k, j' \neq k'$,

$$\begin{aligned}\mathbb{E}(X_j + X_k)^{2p} &= O(n^{-p}); \\ \text{Cov}((X_j + X_k)^{2p+1}, (X_{j'} + X_{k'})^{2q+1}) &= \begin{cases} O(n^{-p-q-1}), & \text{if } \{j, k\} \cap \{j', k'\} \neq \emptyset \\ O(n^{-p-q-2}), & \text{otherwise;} \end{cases} \\ \text{Cov}((X_j + X_k)^{2p}, (X_{j'} + X_{k'})^{2q}) &= \begin{cases} O(n^{-p-q}), & \text{if } \{j, k\} \cap \{j', k'\} \neq \emptyset \\ O(n^{-p-q-2}), & \text{otherwise.} \end{cases}\end{aligned}$$

Each of these is an obvious consequence of Lemma 4.2 except perhaps the last claim. Consider monomials of the form $\mu\mu'$, where μ is a monomial in θ_j, θ_k and μ' is a monomial in $\theta_{j'}, \theta_{k'}$. Pairings of the terms of $\mu\mu'$ which consist of a pairing of the terms of μ together with a pairing of the terms of μ' occur with the same constant in both $\mathbb{E}((X_j + X_k)^{2p}(X_{j'} + X_{k'})^{2q})$ and $\mathbb{E}(X_j + X_k)^{2p} \mathbb{E}(X_{j'} + X_{k'})^{2q}$. Because both μ and μ' are even, any other pairing of the terms of $\mu\mu'$ contains at least two of $\sigma_{jj'}, \sigma_{jk'}, \sigma_{kj'}, \sigma_{kk'}$, so its value is at most $O(n^{-p-q-2})$.

Now we can just apply these bounds to the definition of f_H . It helps to use the fact that for real random variables X_1, \dots, X_m we have $\text{Var}(\sum_{j=1}^m X_j) = \sum_{j,k=1}^m \text{Cov}(X_j, X_k)$. \square

Now we can complete the proof of Theorem 5.2 by applying Theorem 4.4 to estimate $\int_{\Omega} G_{\mathbf{d},H}(\boldsymbol{\theta}) d\boldsymbol{\theta}$. From Remark 4.5 and the norm bound in Lemma 5.4, we can take $\rho_1 = a_2^{-1} \log n$ and $\rho_2 = a_3 \log n$. For $\boldsymbol{\theta} \in T(U_n(\rho_2))$, we have by Taylor's theorem that

$$G_{\mathbf{d},H}(\boldsymbol{\theta}) = e^{-\boldsymbol{\theta}^T A \boldsymbol{\theta} + f_H(\boldsymbol{\theta}) + h(\boldsymbol{\theta})},$$

where $h(\boldsymbol{\theta}) = O(n^{-1/2}(\log n)^5)$.

From the definition of f_H we find for $\boldsymbol{\theta} \in T(U_n(\rho_2))$ that $\partial f_H / \partial \theta_j = O(s_{\max} + (\log n)^2)$ for all j . Similarly, for $j \neq k$, $\partial^2 f_H / \partial \theta_j \partial \theta_k = O(1)$ if $\{j, k\} \in H^+ \cup H^-$ and $O(n^{-1/2} \log n)$ otherwise. Finally, $\partial^2 f_H / \partial \theta_j^2 = O(n^{1/2} \log n)$ for all j . From the last two bounds we have $\|H(f_H, T(U_n(\rho_2)))\|_{\infty} = O(n^{1/2} \log n)$. This gives us a value of $\phi_1 = O(s_{\max} n^{-1/6} \log n)$.

The function g in Theorem 4.4 can be taken to be $\Re f_H$, whose first derivatives are bounded by $O(n^{-1/3} \log n)$ and Hessian by $\|H(\Re f_H, T(U_n(\rho_2)))\|_{\infty} = O(n^{1/6})$. This gives a value of $\phi_2 = O(n^{-1/2+\varepsilon})$.

We now find that all the conditions of Theorem 4.4 are satisfied. Apply Lemma 5.5 using $\mathbb{V}f_H = \text{Var} \Re f_H - \text{Var} \Im f_H = \text{Var} \Re f_H - \mathbb{E}(\Im f_H)^2$, since $\mathbb{E} \Im f_H = 0$. Finally, apply Lemma 5.3 with $k = 1$. To estimate $\int_{\Omega} |G_{(\emptyset, \emptyset)}|$ use the same arguments as above using $\Re f_{(\emptyset, \emptyset)}(\boldsymbol{\theta})$ in place of $f_H(\boldsymbol{\theta})$. This gives an added error term that fits into K . Note that

our conditions on s_{\max} allow for $K = -1$, but even in that case the theorem is valid and gives a useful upper bound. Finally, we can perform the division $N_H(\mathbf{d})/N_{(\emptyset, \emptyset)}(\mathbf{d})$ to obtain $P_H(\mathbf{d})$, noting that for the denominator the error term K is $o(1)$. \square

As we will demonstrate in Subsection 5.3, for obtaining concentration results it is worth noting that the same method gives an upper bound for larger subgraphs.

Theorem 5.6. *Let \mathbf{d} be δ -tame for some $\delta > 0$. Define $\{\lambda_{jk}\}, S, s_{\max}$ as above, and suppose that $s_{\max} \leq b_1 n^{2/3}/(\log n)^2$ and $S \leq b_2 n$ for some constants $b_1, b_2 > 0$. Then there is $c' = c'(\delta, b_1, b_2)$ such that*

$$P_H(\mathbf{d}) \leq c' \prod_{\{j,k\} \in H^+} \lambda_{jk} \prod_{\{j,k\} \in H^-} (1 - \lambda_{jk}).$$

Proof. The proof is the same as for Theorem 5.2 except that we bound $N_H(\mathbf{d})$ by using $|G_{\mathbf{d}, H}(\boldsymbol{\theta})|$ in place of $G_{\mathbf{d}, H}(\boldsymbol{\theta})$. This corresponds to dropping the imaginary parts of f_H .

If $g(\theta) = \Re f_H(\theta)$ and $\boldsymbol{\theta} \in T(U_n(\rho_2))$, then $\mathbb{E} g(\mathbf{X}) = O(1)$, $\text{Var } g(\mathbf{X}) = O(n^{-1/3})$, $|g_j| = O((s_{\max} + (\log n)^2) \log n / n^{1/2})$, and $\|H(g, T(U_n(\rho_2)))\|_{\infty} = O(s_{\max} + (\log n)^2)$, where in each case the implied constant depends only on δ, b_1, b_2 . Applying Theorem 4.4 as before gives the theorem. \square

5.2 Bipartite graphs

Define $V_1, V_2, n_1, n_2, \tilde{E}$ as before. To keep the notation parallel to the notation in the previous section, we will assume that \tilde{E}, H^+, H^- are disjoint.

This case is not covered by the previous subsection since the set of forbidden edges is too big for Theorems 5.2 and 5.6. Nevertheless, we will derive similar results by using formula (5.4) with the radii chosen in such a way that the contours pass through the saddle point for $H = (\emptyset, \tilde{E})$. Accordingly, let $\tilde{\lambda}_{jk}$ be the value of p_{jk} in the solution of (5.3) subject to (5.6) in the case $H = (\emptyset, \tilde{E})$.

Define $\tilde{N}_H(\mathbf{d}) = N_{(H^+, H^- \cup \tilde{E})}(\mathbf{d})$, $\tilde{P}_H(\mathbf{d}) = P_{(H^+, H^- \cup \tilde{E})}(\mathbf{d})$, and $\tilde{C}_H(\mathbf{d}) = C_{(H^+, H^- \cup \tilde{E})}(\mathbf{d})$. Thus $\tilde{N}_H(\mathbf{d})$ is the number of bipartite graphs with degrees \mathbf{d} , on (V_1, V_2) that contain H^+ and are disjoint from H^- , and $\tilde{P}_H(\mathbf{d})$ is the fraction of such graphs among all bipartite graphs on (V_1, V_2) with degrees \mathbf{d} .

Define $\tilde{G}_{\mathbf{d}, H}(\boldsymbol{\theta}) = G_{\mathbf{d}, (H^+, H^- \cup \tilde{E})}(\boldsymbol{\theta})$ and $\tilde{f}_H(\boldsymbol{\theta}) = f_{(H^+, H^- \cup \tilde{E})}(\boldsymbol{\theta})$ as in (5.5) and (5.7), but using $\{\tilde{\lambda}_{jk}\}$ instead of $\{\lambda_{jk}\}$.

With a tiny adjustment, we adopt from Barvinok and Hartigan [5] conditions on \mathbf{d} that we call δ -bitame for $\delta > 0$: $\delta \leq \tilde{\lambda}_{jk} \leq 1 - \delta$ for all $\{j, k\} \notin \tilde{E}$ and $n_1, n_2 \geq \delta n$. We provide a sufficient condition similar to Lemma 5.1.

Lemma 5.7. *Let p, q be real numbers such that $0 < q^2 < p \leq q < 1$. Then for any degree sequence d_1, \dots, d_n such that $\sum_{j \in V_1} d_j = \sum_{j \in V_2} d_j$ and*

$$pn_2 \leq d_j \leq qn_2 \text{ for } j \in V_1, \quad pn_1 \leq d_j \leq qn_1 \text{ for } j \in V_2,$$

the solution $\{\tilde{\lambda}_{jk}\}$ defined above exists and $\delta < \tilde{\lambda}_{jk} < 1 - \delta$ for all $\{j, k\} \notin \tilde{E}$, where $\delta > 0$ depends only on p, q .

Proof. Without loss of generality we can assume that $d_1 \geq \dots \geq d_{n_1}$. To prove the existence of the solution $\{\tilde{\lambda}_{jk}\}$ it will suffice to show that all Gale-Ryser inequalities are strict [40]; i.e., for any $1 \leq k < n_1$,

$$\sum_{j=1}^k d_j < \sum_{j \in V_2} \min\{d_j, k\}.$$

If $qn_1 < k < n_1$ then $\sum_{j \in V_2} \min\{d_j, k\} = \sum_{j \in V_2} d_j > \sum_{j=1}^k d_j$. For $k < pn_1$ we get that $\sum_{j \in V_2} \min\{d_j, k\} = kn_2 > kqn_2 \geq \sum_{j=1}^k d_j$. For the remaining case, when $pn_1 \leq k \leq qn_1$, observe that $\sum_{j \in V_2} \min\{d_j, k\} \geq pn_1 n_2 > q^2 n_1 n_2 \geq kqn_2$.

If $\{\beta_j\}$ are the parameters in (5.3) corresponding to $\{\tilde{\lambda}_{jk}\}$, and c is a constant, recall that $\beta_1 - c, \dots, \beta_{n_1} - c, \beta_{n_1+1} + c, \dots, \beta_n + c$ is also a solution. By choice of c , we can assume for some $\gamma \in [0, 1]$ that

$$|V_1^+| \geq \gamma n_1, \quad |V_1^-| \geq (1 - \gamma)n_1, \quad |V_2^+| \geq \gamma n_2, \quad |V_2^-| \geq (1 - \gamma)n_2, \quad (5.9)$$

where $V_t^\pm = \{j \in V_t \mid \pm \beta_j \geq 0\}$. Recalling that for $\{j, k\} \notin \tilde{E}$

$$\tilde{\lambda}_{jk} = \frac{e^{\beta_j + \beta_k}}{1 + e^{\beta_j + \beta_k}},$$

it is sufficient to show that $|\beta_j| \leq b$, $j = 1, \dots, n$ for some $b = b(p, q) > 0$.

Define $a = \max_{j \in V_1} \beta_j$ and $b = \min_{j \in V_2} \beta_j$. Without loss of generality, we can assume that $a = \beta_1$ and $b = \beta_n$. Note that

$$d_1 = \sum_{j \in V_2} \tilde{\lambda}_{1j} \geq n_2 \frac{e^{a+b}}{1 + e^{a+b}}, \quad d_n = \sum_{j \in V_1} \tilde{\lambda}_{jn} \leq n_1 \frac{e^{a+b}}{1 + e^{a+b}}.$$

By assumption $d_1 \leq qn_2$ and $d_n \geq pn_1$, therefore

$$p \leq \frac{e^{a+b}}{1 + e^{a+b}} \leq q \implies \log \frac{p}{1-p} \leq a+b \leq \log \frac{q}{1-q}. \quad (5.10)$$

Using (5.9), we find also that

$$d_1 = \sum_{j \in V_2} \tilde{\lambda}_{1j} \geq \gamma n_2 \frac{e^a}{1+e^a}, \quad d_n = \sum_{j \in V_1} \tilde{\lambda}_{jn} \leq \gamma n_2 \frac{e^{a+b}}{1+e^{a+b}} + (1-\gamma) n_2 \frac{e^b}{1+e^b},$$

which gives us $q \geq \gamma \frac{e^a}{1+e^a}$ and $p \leq \gamma \frac{e^{a+b}}{1+e^{a+b}} + (1-\gamma) \frac{e^b}{1+e^b} \leq \gamma q + (1-\gamma) \frac{e^b}{1+e^b}$. If $\gamma \geq (p+q^2)/2q$ then the first inequality implies $\frac{e^a}{1+e^a} \leq 2q^2/(p+q^2)$. Otherwise, from the second inequality we get that $\frac{e^b}{1+e^b} \geq q(p-q^2)/(2q-p-q^2)$. Using (5.10), we get in the both cases that

$$\max_{j \in V_1} \beta_j = a \leq b \quad \text{and} \quad \min_{j \in V_2} \beta_j = b \geq -b \quad \text{for some } b = b(p, q) > 0.$$

In order to get the missing reverse bounds and to complete the proof, we just need to swap the roles of subsets V_1, V_2 . \square

Define the $n \times n$ symmetric matrix \tilde{A} by

$$\boldsymbol{\theta}^T \tilde{A} \boldsymbol{\theta} = \frac{1}{2} \sum_{\{j,k\} \notin \tilde{E}} \tilde{\lambda}_{jk} (1 - \tilde{\lambda}_{jk}) (\theta_j + \theta_k)^2.$$

For each j , let s_j be the number of times vertex j occurs in $H^+ \cup H^-$ and define $s_{\max} = \max_{j=1}^n s_j$, $S = \frac{1}{2} \sum_{j=1}^n s_j$ and $S_2 = \sum_{j=1}^n s_j^2$. Differently from the matrix A in the previous subsection, \tilde{A} has a zero eigenvalue. Let $\mathbf{w} = (w_1, \dots, w_n)^T$ be defined by $w_j = (-1)^m$ if $j \in V_m$ for $m = 1, 2$. Note that $\ker \tilde{A} = \langle \mathbf{w} \rangle$ and $\tilde{f}(\boldsymbol{\theta} + t\mathbf{w}) = \tilde{f}(\boldsymbol{\theta})$ for any $t \in \mathbb{R}$ and $\boldsymbol{\theta} \in \mathbb{R}^n$.

Theorem 5.8. *Let \mathbf{d} be δ -bitame for some $\delta > 0$. Define $\{\tilde{\lambda}_{jk}\}, \tilde{A}, s_{\max}, S_2, \tilde{f}_H, \mathbf{w}$ as above, and suppose that $s_{\max} \leq c_1 n^{1/6}$ and $S_2 \leq c_2 n$ for some constants c_1, c_2 . Let $\tilde{\mathbf{X}}$ be a random variable with the normal density $\pi^{-n/2} |\tilde{A} + \mathbf{w}\mathbf{w}^T|^{-1/2} e^{-\mathbf{x}^T (\tilde{A} + \mathbf{w}\mathbf{w}^T) \mathbf{x}}$. Then, for any $\varepsilon > 0$, there is a constant $\tilde{c} = \tilde{c}(\delta, \varepsilon, c_1, c_2)$ such that*

$$N_H(\mathbf{d}) = 2\pi^{(n+1)/2} n \tilde{C}_{\mathbf{d}, H} |\tilde{A} + \mathbf{w}\mathbf{w}^T|^{-1/2} e^{\mathbb{E} \Re \tilde{f}_H(\tilde{\mathbf{X}}) - \frac{1}{2} \mathbb{E} (\Im \tilde{f}_H(\tilde{\mathbf{X}}))^2} (1 + \tilde{K}), \quad (5.11)$$

where $|\tilde{K}| \leq e^{\tilde{c}(1+s_{\max}^3)n^{-1/2+\varepsilon}} - 1$. Moreover,

$$\tilde{P}_H(\mathbf{d}) = (1 + \tilde{K}') \prod_{\{j,k\} \in H^+} \tilde{\lambda}_{jk} \prod_{\{j,k\} \in H^-} (1 - \tilde{\lambda}_{jk})$$

where $|\tilde{K}'| \leq e^{\tilde{c}S_2/n + \tilde{c}(1+s_{\max}^3)n^{-1/2+\varepsilon}} - 1$.

Using Corollary 4.7 one can note that (5.11) in the case of $H = (\emptyset, \emptyset)$ (with a different error term) matches [5, formula (2.5.4)]. The formula for $\tilde{P}_H(\mathbf{d})$ is given more precisely in [15], but only for the near-semiregular degree sequences considered there.

Proof. We start from formula (5.4). Since $\tilde{G}_{\mathbf{d},H}(\boldsymbol{\theta} + t\mathbf{w}) = \tilde{G}_{\mathbf{d},H}(\boldsymbol{\theta})$ for any $t \in \mathbb{R}$ and $\boldsymbol{\theta} \in \mathbb{R}^n$ we can fix $\theta_n = 0$ and multiply by 2π

$$\tilde{N}_H(\mathbf{d}) = 2\pi \tilde{C}_{\mathbf{d},H} \int_{U_{n-1}(\pi)} \tilde{G}_{\mathbf{d},H}(\boldsymbol{\theta}) d\boldsymbol{\theta}',$$

where $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\theta}') = (\theta'_1, \dots, \theta'_{n-1}, 0)$. Let $\Omega = U_n(\log n/n^{1/2})$ and $L = \{\boldsymbol{\theta} \in \mathbb{R}^n \mid \theta_n = 0\}$.

Lemma 5.9. *For any $k > 0$,*

$$\int_{U_{n-1}(\pi)} \tilde{G}_{\mathbf{d},H}(\boldsymbol{\theta}) d\boldsymbol{\theta}' = \int_{\Omega \cap L} \tilde{G}_{\mathbf{d},H}(\boldsymbol{\theta}) d\boldsymbol{\theta}' + O(n^{-k}) \int_{\Omega \cap L} |\tilde{G}_{\mathbf{d},(\emptyset,\emptyset)}(\boldsymbol{\theta})| d\boldsymbol{\theta}'$$

Proof. This follows from [5, p. 340] in the same way that Lemma 5.3 follows from [5, Thm. 8.1]. Note that Barvinok and Hartigan do not actually provide a proof, but we agree with them that there is a proof parallel to that of their Theorem 8.1. \square

Define matrices Q, W, P, R by

$$\begin{aligned} Q\mathbf{x} &= \mathbf{x} - x_n \mathbf{w}, & W\mathbf{x} &= \frac{1}{\sqrt{n}} \mathbf{w} \mathbf{w}^T \mathbf{x}, \\ P\mathbf{x} &= \mathbf{x} - \frac{1}{n} \mathbf{w} \mathbf{w}^T \mathbf{x}, & R\mathbf{x} &= \frac{1}{\sqrt{n}} \mathbf{x}. \end{aligned}$$

Applying Lemma 4.6 with $\rho = \log n$, we find that

$$\begin{aligned} \int_{\Omega \cap L} \tilde{G}_{\mathbf{d},H}(\boldsymbol{\theta}) d\boldsymbol{\theta}' &= (1 + O(n^{-\log n})) \pi^{-1/2} n \int_{\Omega_\rho} \tilde{G}_{\mathbf{d},H}(\boldsymbol{\theta}) e^{-\boldsymbol{\theta}^T \mathbf{w} \mathbf{w}^T \boldsymbol{\theta}} d\boldsymbol{\theta}, \\ \int_{\Omega \cap L} |\tilde{G}_{\mathbf{d},(\emptyset,\emptyset)}(\boldsymbol{\theta})| d\boldsymbol{\theta} &= (1 + O(n^{-\log n})) \pi^{-1/2} n \int_{\Omega_\rho} |\tilde{G}_{\mathbf{d},(\emptyset,\emptyset)}(\boldsymbol{\theta})| e^{-\boldsymbol{\theta}^T \mathbf{w} \mathbf{w}^T \boldsymbol{\theta}} d\boldsymbol{\theta}, \end{aligned}$$

and also that $U_n(\frac{1}{2} \log n/n^{1/2}) \subseteq \Omega_\rho \subseteq U_n(3 \log n/n^{1/2})$.

We continue the proof of Theorem 5.8 with a sequence of lemmas.

Lemma 5.10. *Let D be the diagonal matrix with the same diagonal as \tilde{A} . Then for some constant a_1 we have $\|(\tilde{A} + \mathbf{w} \mathbf{w}^T)^{-1} - D^{-1}\|_{\max} \leq a_1 n^{-2}$. Furthermore, there exists a matrix T with $T^T(\tilde{A} + \mathbf{w} \mathbf{w}^T)T = I$ and some constants a_2, a_3 such that $\|T\|_1, \|T\|_\infty \leq a_2 n^{-1/2}$ and $\|T^{-1}\|_\infty \leq a_3 n^{1/2}$.*

Proof. From the definition of \tilde{A} we have $\|\tilde{A} - D\|_{\max} \leq \frac{1}{8}$. Also, for any \mathbf{x} such that $\mathbf{w}^T \mathbf{x} = 0$ we have

$$\mathbf{x}^T (\tilde{A} + \mathbf{w} \mathbf{w}^T) \mathbf{x} = \mathbf{x}^T \tilde{A} \mathbf{x} \geq \frac{1}{2} \delta (1 - \delta) \sum_{\{j,k\} \in \tilde{E}} (x_j + x_k)^2 \geq \frac{1}{2} \delta (1 - \delta) \delta n \mathbf{x}^T \mathbf{x},$$

where we used the fact that all eigenvalues with exception of one zero (which corresponds to \mathbf{w}) of the quadratic form $\sum_{\{j,k\} \notin \tilde{E}} (x_j + x_k)^2$ are at least $\min\{|V_1|, |V_2|\} \geq \delta n$. We note also that $\|\mathbf{w}^T \mathbf{w}\|_{\max} = 1$ and for any $\mathbf{x} = t\mathbf{w}$

$$\mathbf{x}^T (\tilde{A} + \mathbf{w}^T \mathbf{w}) \mathbf{x} = \mathbf{x}^T \mathbf{w}^T \mathbf{w} \mathbf{x} = n \mathbf{x}^T \mathbf{x}.$$

Taking into account the following inequalities:

$$\begin{aligned} \max D_{jj} &\leq \frac{1}{2} \delta (1 - \delta) \max\{|V_1|, |V_2|\} \leq \frac{1}{8} (1 - \delta) n, \\ \min D_{jj} &\geq \frac{1}{2} \delta (1 - \delta) \min\{|V_1|, |V_2|\} \geq \frac{1}{2} \delta^2 (1 - \delta) n, \end{aligned}$$

we finish the proof by applying Lemma 4.9 with $r = 9/(4\delta^2(1 - \delta))$ and $\gamma = 4\delta^2$. \square

Lemma 5.11. *We have*

$$\begin{aligned} \mathbb{E} \tilde{f}_H(\tilde{\mathbf{X}}) &= \mathbb{E} \tilde{f}_{(\emptyset, \emptyset)}(\tilde{\mathbf{X}}) + O(S/n) = O(1), \\ \text{Var } \Re \tilde{f}_H(\tilde{\mathbf{X}}) &= \text{Var } \Re \tilde{f}_{(\emptyset, \emptyset)}(\tilde{\mathbf{X}}) + O(S_2/n^2) = O(1/n), \quad \text{and} \\ \text{Var } \Im \tilde{f}_H(\tilde{\mathbf{X}}) &= \text{Var } \Im \tilde{f}_{(\emptyset, \emptyset)}(\tilde{\mathbf{X}}) + O(S_2/n) = O(1). \end{aligned}$$

Lemma 5.11 is proved in precisely the same way as Lemma 5.5.

In order to estimate $\int_{\Omega_\rho} \tilde{G}_{\mathbf{d}, H}(\boldsymbol{\theta}) e^{-\boldsymbol{\theta}^T \mathbf{w} \mathbf{w}^T \boldsymbol{\theta}} d\boldsymbol{\theta}$ and $\int_{\Omega_\rho} |\tilde{G}_{\mathbf{d}, (\emptyset, \emptyset)}(\boldsymbol{\theta})| e^{-\boldsymbol{\theta}^T \mathbf{w} \mathbf{w}^T \boldsymbol{\theta}} d\boldsymbol{\theta}$ we apply Theorem 4.4. From Remark 4.5 and the norm bound in Lemma 5.10, we can take $\rho_1 = \frac{1}{2} a_2^{-1} \log n$ and $\rho_2 = 3a_3 \log n$. For $\boldsymbol{\theta} \in T(U_n(\rho_2))$, we have by Taylor's theorem that

$$\tilde{G}_{\mathbf{d}, H}(\boldsymbol{\theta}) e^{-\boldsymbol{\theta}^T \mathbf{w} \mathbf{w}^T \boldsymbol{\theta}} = e^{-\boldsymbol{\theta}^T (\tilde{A} + \mathbf{w} \mathbf{w}^T) \boldsymbol{\theta} + \tilde{f}_H(\boldsymbol{\theta}) + \tilde{h}(\boldsymbol{\theta})},$$

where $\tilde{h}(\boldsymbol{\theta}) = O(n^{-1/2}(\log n)^5)$. Now we finish the proof in complete analogy with the proof of Theorem 5.2. \square

The same argument gives us also the analog of Theorem 5.6 for bipartite case which will be useful for obtaining concentration results.

Theorem 5.12. *Let \mathbf{d} be δ -bitame for some $\delta > 0$. Define $\{\tilde{\lambda}_{jk}\}$, S , s_{\max} as above, and suppose that $s_{\max} \leq b_1 n^{2/3}/(\log n)^2$ and $S \leq b_2 n$ for some constants $b_1, b_2 > 0$. Then there is $\tilde{c}' = \tilde{c}'(\delta, b_1, b_2)$ such that*

$$\tilde{P}_H(\mathbf{d}) \leq \tilde{c}' \prod_{\{j,k\} \in H^+} \tilde{\lambda}_{jk} \prod_{\{j,k\} \in H^-} (1 - \tilde{\lambda}_{jk}).$$

Remark 5.13. Theorems 5.2 and 5.8 are less general than our techniques allow, due to the choices that we made here for the purpose of keeping our example simple. We restricted ourselves to δ -tame and δ -bitame degree sequences so that we could adopt Lemmas 5.3

and 5.9 from [5]. More significantly, we used the saddle point of f for f_H as well, which simplifies the calculation a lot at the expense of restricting H far more than necessary. In a follow-up paper, we will show how to estimate $N_H(\mathbf{d})$ whenever the quadratic form

$$\sum_{jk \notin H^+ \cup H^-} \lambda_{jk}(1 - \lambda_{jk})(\theta_j + \theta_k)^2$$

has not too many zero eigenvalues and all its non-zero eigenvalues are at least δn , where $\delta > 0$ may be constant or slowly decreasing.

5.3 Concentration near the β -model

For a given degree sequence \mathbf{d} and pair of vertices $j \neq k$ (in bipartite case, for $\{j, k\} \notin \tilde{E}$), let $\xi_{jk} = \xi_{jk}(\mathbf{d})$ be the indicator variable for $\{j, k\}$ being an edge in a uniformly random graph (or, alternatively, a uniformly random bipartite graph with partite sets V_1, V_2) with degree sequence \mathbf{d} . Let $\{\hat{\xi}_{jk}\}$ be independent Bernoulli variables with $\text{Prob}(\hat{\xi}_{jk} = 1) = \lambda_{jk}$ for all j, k (or, in the bipartite case, $\text{Prob}(\hat{\xi}_{jk} = 1) = \tilde{\lambda}_{jk}$ for pairs $\{j, k\} \notin \tilde{E}$). Note that $\{\hat{\xi}_{jk}\}$ is just the β -model.

Theorems 5.2 and 5.8 show that $\{\xi_{jk}\}$ and $\{\hat{\xi}_{jk}\}$ are point-wise almost identical at small scales. Now we explore their relationship at large scales. Let Y be a set of vertex pairs (disjoint from \tilde{E} in the bipartite case). Define $X = X(Y, \mathbf{d}) = \sum_{jk \in Y} \xi_{jk}$ and $\hat{X} = \hat{X}(Y, \mathbf{d}) = \sum_{jk \in Y} \hat{\xi}_{jk}$. From Theorems 5.2 and 5.8 we have that $\mathbb{E} \hat{X}^t \sim \mathbb{E} X^t$ for $t = O(n^{1/6-\epsilon})$, but this is not sufficient to estimate $\text{Var } \hat{X}$.

Barvinok [3], in the bipartite case under conditions more general than δ -bitameness, and Barvinok and Hartigan [5] in the general case under δ -tameness, show that for, $|Y| \geq \delta n^2$,

$$(1 - \delta n^{-1/2} \log n) \mathbb{E} \hat{X} \leq X \leq (1 + \delta n^{-1/2} \log n) \mathbb{E} \hat{X} \quad (5.12)$$

with probability $1 - n^{-\Omega(n)}$. In the case of near-regular degree sequences, McKay [31] proved a weaker concentration of X near $\mathbb{E} \hat{X}$ whenever $|Y| \rightarrow \infty$. Note that (5.12) starts to “bite” at around $|Y|^{3/4}$ from the mean. Since the variance of \hat{X} has the same order as the expectation of \hat{X} for all Y , it seems likely that a concentration inequality that bites at around $|Y|^{1/2}$ from the mean is the best that can be hoped for without specifying more structure for Y . Here we prove such concentration in both the general and bipartite cases, starting with a lemma that bounds the moments of X in terms of the moments of \hat{X} .

Lemma 5.14. *Let the assumptions of Theorem 5.15 hold. Then for any $b > 0$, there is a constant $\hat{c} = \hat{c}(\delta, b) > 0$ such that $\mathbb{E} X^m \leq \hat{c} \mathbb{E} \hat{X}^m$ for all integers m with $0 \leq m \leq b|Y|^{1/2}n^{1/6}/(\log n)^3$.*

Proof. For $1 \leq t \leq m$, let

$$X_t = \sum_{W \subseteq Y: |W|=t} \prod_{jk \in W} \xi_{jk}, \quad \text{and} \quad \hat{X}_t = \sum_{W \subseteq Y: |W|=t} \prod_{jk \in W} \hat{\xi}_{jk}.$$

Since these are indicator variables, we have

$$X^m = \sum_{t=1}^m t! \left\{ \begin{matrix} m \\ t \end{matrix} \right\} X_t \quad \text{and} \quad \hat{X}^m = \sum_{t=1}^m t! \left\{ \begin{matrix} m \\ t \end{matrix} \right\} \hat{X}_t,$$

where $\left\{ \begin{matrix} m \\ t \end{matrix} \right\}$ is the Stirling number of the second kind. It follows that the assertion will be true if $\mathbb{E} X_t \leq \hat{c} \mathbb{E} \hat{X}_t$ for $1 \leq t \leq m$, where \hat{c} is a constant depending only on b and δ . Due to Theorem 5.6, we immediately get this bound if $m \leq b_1 n^{2/3}/(\log n)^2$ (and, consequently, if $|Y| \leq (b_1/b)^2 n (\log n)^2$) for any fixed $b_1 > 0$. For greater values of m and $|Y|$, it requires additional consideration.

Any subset $W \subseteq Y$ induces a graph on n vertices. Let w_j denote the degree of j in this graph. We refer to a vertex j as a W -full vertex if $w_j > \lfloor n^{2/3}/(\log n)^2 \rfloor$ and a pair $jk \in Y$ as a W -critical pair if at least one of the vertices j, k is W -full. Define $\eta(W) = \sum_{j=1}^n \max\{0, w_j - \lfloor n^{2/3}/(\log n)^2 \rfloor\}$. Since a set satisfying Theorem 5.6 is obtained by removing at most $\eta(W)$ elements from W , we have that

$$\mathbb{E} \left(\prod_{jk \in W} \xi_{jk} \right) \leq p(W), \quad \text{where} \quad p(W) = c' \delta^{-\eta(W)} \prod_{jk \in W} \lambda_{jk},$$

where c' is the constant from Theorem 5.6. Consequently,

$$\mathbb{E} X_t \leq c' \sum_{W \subseteq Y: |W|=t} p(W).$$

We now apply Lemma 6.4, stated in the Appendix. Define a digraph D whose vertices are the t -subsets of Y . The ordered pair (W, W') is an edge of D if $W - W'$ consists of one element, which is W -critical. Define $s, \alpha : E(D) \rightarrow \mathbb{R}$ by

$$s(W, W') = \frac{p(W)p(W')}{\sum_{W'': (W'', W') \in E(D)} p(W'')} \quad \text{and} \quad \alpha(W, W') = \frac{\sum_{W'': (W'', W') \in E(D)} p(W'')}{\sum_{W'': (W, W'') \in E(D)} p(W'')}.$$

It is routine to check that the conditions of Lemma 6.4 are satisfied, with Z being the set of vertices W with $\eta(W) = 0$, provided we have $\alpha(W, W') < 1$ for every edge.

Given W with $\eta(W) > 0$, we can choose a W -critical pair belonging to W in at least $n^{2/3}/(\log n)^2$ ways, and then we can replace it by some element of $Y - W$ in at least $|Y| - t$

ways. Thus, the out-degree of W is at least $n^{2/3}(|Y| - t)/(\log n)^2$. Alternatively, given W' , we can choose an element $jk \in W'$ in t ways, choose a vertex presented in at least $\lfloor n^{2/3}/(\log n)^2 \rfloor$ pairs of W' (i.e. W' -full or almost W' -full) in at most $t/\lfloor n^{2/3}/(\log n)^2 \rfloor$ ways and replace jk by a pair containing that vertex in at most n ways. So the in-degree of W' is at most $t^2 n / \lfloor n^{2/3}/(\log n)^2 \rfloor$. Finally, an in-neighbour W_1 of W' differs in at most 3 elements from an out-neighbour W_2 of W , so $|\eta(W_1) - \eta(W_2)| \leq 6$. Since $t \leq m$, we find that $\alpha(W, W') \leq \hat{\alpha} = 2b^2\delta^{-6}/(\log n)^2 < \frac{1}{2}$ for large enough n .

Consequently, since $p(W) = c' \prod_{jk \in W} \lambda_{jk}$ when $\eta(W) = 0$, Lemma 6.4 tells us that

$$\mathbb{E} X_t \leq \frac{1 - \hat{\alpha}}{1 - 2\hat{\alpha}} \sum_{W \subseteq Y: \eta(W)=0} p(W) \leq c' \frac{1 - \hat{\alpha}}{1 - 2\hat{\alpha}} \sum_{W \subseteq Y: |W|=t} \prod_{jk \in W} \lambda_{jk} = c' \frac{1 - \hat{\alpha}}{1 - 2\hat{\alpha}} \mathbb{E} \hat{X}_t.$$

This completes the proof. \square

Theorem 5.15. *Suppose \mathbf{d} is δ -tame (or δ -bitame) for some $\delta > 0$. Let Y be a set of vertex pairs (disjoint from \tilde{E} in the bipartite case). Then for any $\gamma > 0$*

$$P(|X - \mathbb{E} \hat{X}| < \gamma |Y|^{1/2}) \geq 1 - \check{c} e^{-2\gamma \min\{\gamma, n^{1/6}(\log n)^{-3}\}},$$

where the constant $\check{c} > 0$ depends only on δ .

Proof. The proofs of the general and bipartite cases are the same; we will use the notation of the general case. Let $p > 0$ be such that $p|Y| = \mathbb{E} \hat{X} = \sum_{jk \in Y} \lambda_{jk}$. Hoeffding's Lemma (see [17] and Lemma 2.5) gives us for any $t > 0$

$$\mathbb{E} e^{t\hat{X}} \leq e^{tp|Y| + \frac{1}{8}t^2|Y|}.$$

Using Lemmas 5.14 and 6.3, we find that for $t \leq \frac{4n^{1/6}}{(\log n)^3} |Y|^{-1/2}$

$$\mathbb{E} e^{tX} \leq \frac{4}{3} \sum_{k=0}^{\lfloor \frac{16n^{1/6}|Y|^{1/2}}{(\log n)^3} \rfloor} \frac{t^k \mathbb{E} X^k}{k!} \leq \frac{4}{3} \hat{c} \mathbb{E} e^{t\hat{X}} \leq \frac{4}{3} \hat{c} e^{tp|Y| + \frac{1}{8}t^2|Y|}.$$

Taking $t = 4|Y|^{-1/2} \min\{\frac{n^{1/6}}{(\log n)^3}, \gamma\}$ and using Markov's inequality for e^{tX} , we obtain that

$$P(X \geq p|Y| + \gamma|Y|^{1/2}) \leq \frac{\mathbb{E} e^{tX}}{e^{tp|Y| + t\gamma|Y|^{1/2}}} \leq \frac{4}{3} \hat{c} e^{-t\gamma|Y|^{1/2} + \frac{1}{8}t^2|Y|} \leq \frac{4}{3} \hat{c} e^{-2\gamma \min\{\gamma, \frac{n^{1/2}}{(\log n)^3}\}}.$$

To complete the proof we apply the same arguments for the complement degree sequence $\mathbf{d}^c = (n - 1 - d_1, \dots, n - 1 - d_n)$ which is also δ -tame with $\lambda_{jk}^c = 1 - \lambda_{jk}$ \square

6 Appendix

Here we give proofs of some technical results that were used in the proofs.

Lemma 6.1. *If $z_1, z_2 \in \mathbb{C}$ satisfy $|z_1| \leq \alpha$ and $|z_2| \leq \beta$, then*

$$\begin{aligned} |e^{z_1} - e^{z_1^2/2} - z_1| &\leq e^{\frac{1}{6}\alpha^3 + \frac{1}{8}\alpha^4} - 1, \\ |z_1(e^{z_2} - z_2 - 1)| &\leq e^{\frac{1}{8}\beta^2} + e^{\frac{1}{3}\alpha\beta + \frac{1}{4}\beta^2 + \frac{1}{4}\alpha^4} - \frac{1}{3}\alpha\beta - 2. \end{aligned}$$

Proof. From the signs of the Taylor coefficients of $e^{z_1} - e^{z_1^2/2} - z_1$ we see that the left side of the first inequality is largest when $z_1 = -\alpha$. This means we only need to prove $\phi(\alpha) \geq 0$ for $\alpha \geq 0$, where

$$\phi(\alpha) = e^{\alpha^3/6 + \alpha^4/8} - e^{\alpha^2/2} + e^{-\alpha} + \alpha - 1.$$

It is clear that $\phi(\alpha) > 0$ for $\alpha > \frac{3}{2}$, since in that case $\frac{1}{6}\alpha^3 + \frac{1}{8}\alpha^4 > \alpha^2/2$ and $e^{-\alpha} > 1 - \alpha$. For $0 \leq \alpha \leq \frac{3}{2}$ we can apply $e^x \leq 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{18}x^4$ for $0 \leq x \leq \frac{6}{5}$, $e^x \leq 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3$ for $x \geq 0$, and $e^{-x} \geq \sum_{i=0}^5 \frac{1}{i!}(-x)^i$ for $x \geq 0$. This gives us a polynomial of degree 12 that is nonnegative for all $\alpha \geq 0$ and bounds $\phi(\alpha)$ from below.

For the second inequality, the worst case is obviously $z_1 = \alpha, z_2 = \beta$, so we just need to prove that $\varphi(\alpha, \beta) \geq 0$ for $\alpha, \beta \geq 0$, where

$$\varphi(\alpha, \beta) = e^{\beta^2/8} + e^{\alpha\beta/3 + \beta^2/4 + \alpha^4/4} - \alpha e^{\beta} + \frac{2}{3}\alpha\beta + \alpha - 2.$$

For $0 \leq \alpha \leq 1$, we have $\varphi(\alpha, \beta) \geq e^{\beta^2/8} + e^{\beta^2/4} - e^{\beta} - 2$, which is positive when $\beta \geq 4$. For $\alpha > 1$, note that $e^{\alpha^4/4} > \alpha$, so we have $\varphi(\alpha, \beta) \geq (e^{\beta^2/4} - e^{\beta})\alpha + e^{\beta^2/8} - 2$, and both coefficients are positive for $\beta > 4$. Thus, $\varphi(\alpha, \beta) \geq 0$ for $\alpha \geq 0, \beta \geq 4$.

For $0 \leq \beta \leq 4$, $\varphi(\alpha, \beta) \geq e^{\alpha^4/4} - \alpha e^{\beta} + \alpha - 1$, which is positive when $\alpha \geq 3$.

We are left with the rectangle $R = \{(\alpha, \beta) \mid 0 \leq \alpha \leq 3, 0 \leq \beta \leq 4\}$. A polynomial $\bar{\varphi}(\alpha, \beta)$ such that $\bar{\varphi} \leq \varphi$ on R is obtained using the bounds $e^x \geq 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3$ for $x \geq 0$ and $e^x \leq 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}x^4 + \frac{1}{50}x^5$ for $0 \leq x \leq 4$. We will show that $\bar{\varphi}$ is nonnegative on R .

Using Sturm sequences, we find that $\varphi(\alpha, \beta) > 0$ everywhere on the boundary of R except at the point $(0, 0)$, where it is zero. As $(\alpha, \beta) \rightarrow (0, 0)$, $\bar{\varphi}(\alpha, \beta) = (1 + o(1))(\frac{3}{8}\beta^2 + \frac{1}{4}\alpha^4)$, which is positive in some punctured neighbourhood of $(0, 0)$. Therefore, there is some $\varepsilon > 0$ such that $\bar{\varphi}(\alpha, \beta) \geq 0$ for $0 \leq \alpha < \varepsilon, 0 \leq \beta \leq 4$ and $\bar{\varphi}(\alpha, \beta) > 0$ on the boundary of the rectangle $R_\varepsilon = \{(\alpha, \beta) \mid \varepsilon \leq \alpha \leq 3, 0 \leq \beta \leq 4\}$. If $\bar{\varphi}(\alpha, \gamma)$ has a zero inside R_ε , then there is some constant $\alpha' \in (\varepsilon, 3)$ such that the 1-variable polynomial

$\bar{\varphi}(\alpha', \gamma)$ has a multiple zero in $(0, 4)$. However the discriminant of $\bar{\varphi}(\alpha, \gamma)$ with respect to γ is never zero for $0 \leq \alpha \leq 4$. \square

The following lemma is used in combining error terms.

Lemma 6.2. *Let $K_1, K_2, \varepsilon_1, \varepsilon_2 \in \mathbb{C}$ and $\alpha, \delta_1, \delta_2, \delta_3, \delta_4 \geq 0$. Suppose $|K_1| \leq e^{\delta_1} - 1$, $|K_2| \leq e^{\delta_2} - 1$, $|\varepsilon_1| \leq \delta_3$ and $|\varepsilon_2| \leq \delta_4$. Then*

$$(1 + K_1 e^{\alpha + \varepsilon_2})(1 + K_2) e^{\varepsilon_1} = 1 + K e^{\alpha}$$

for some $K \in \mathbb{C}$ with $|K| \leq e^{\delta_1 + \delta_2 + \delta_3 + \delta_4} - 1$.

Proof. For $z \in \mathbb{C}$ it is immediate from the Taylor series that $|e^z| \leq e^{|z|}$ and $|e^z - 1| \leq e^{|z|} - 1$. Bound $|K|$ by bounding $K = e^{-\alpha}((1 + K_1 e^{\alpha + \varepsilon_2})(1 + K_2) e^{\varepsilon_1} - 1)$ term by term, which gives $|K| \leq e^{\delta_1 + \delta_2 + \delta_3 + \delta_4} + e^{-\alpha + \delta_2 + \delta_3} - e^{\delta_2 + \delta_3 + \delta_4} - e^{-\alpha}$. Therefore $e^{\delta_1 + \delta_2 + \delta_3 + \delta_4} - 1 - |K| \geq e^{\delta_2 + \delta_3 + \delta_4} + e^{-\alpha} - e^{-\alpha + \delta_2 + \delta_3} - 1 \geq e^{\delta_2 + \delta_3} + e^{-\alpha} - e^{-\alpha + \delta_2 + \delta_3} - 1$, which is nonnegative by the convexity of the exponential function since both 0 and $-\alpha + \delta_2 + \delta_3$ lie in the interval $[-\alpha, \delta_2 + \delta_3]$ and the average of 0 and $-\alpha + \delta_2 + \delta_3$ lies at the midpoint of the interval. \square

Lemma 6.3. *For any $m \in \mathbb{N}$ and $0 \leq x \leq m/4$,*

$$\sum_{k=0}^{m-1} \frac{x^k}{k!} \leq e^x \leq \frac{4}{3} \sum_{k=0}^{m-1} \frac{x^k}{k!}.$$

Proof. The lower bound is clear. For the upper bound note that by comparing terms

$$e^x \leq \sum_{k=0}^{m-1} \frac{x^k}{k!} + \frac{x^m}{m!} e^x,$$

and so $e^x \leq (1 - \frac{x^m}{m!})^{-1} \sum_{k=1}^{m-1} \frac{x^k}{k!} \leq (1 - \frac{(m/4)^m}{m!})^{-1} \sum_{k=1}^{m-1} \frac{x^k}{k!} \leq \frac{4}{3} \sum_{k=1}^{m-1} \frac{x^k}{k!}$. \square

The following lemma is an immediate corollary of [24, Thm. 3].

Lemma 6.4. *Let D be a finite directed graph, with loops but not parallel edges allowed. Let $p : V(D) \rightarrow \mathbb{R}_{>0}$, $s : E(D) \rightarrow \mathbb{R}_{>0}$ and $\alpha : E(D) \rightarrow (0, 1)$ be functions such that the following inequalities hold.*

$$\begin{aligned} \sum_{w:(vw) \in E(D)} \alpha(vw) s(vw) &\geq p(v), & \text{for } v \in V(D) \text{ not a sink, and} \\ \sum_{v:(vw) \in E(D)} s(vw) &\leq p(w), & \text{for all } w \in V(D). \end{aligned}$$

Let $Z \subseteq V(D)$ be the set of sinks of G . Then

$$\frac{\sum_{v \in V(G) - Z} p(v)}{\sum_{v \in V(G)} p(v)} \leq \frac{\max_{(vw) \in E(D)} \alpha(vw)}{1 - \max_{(vw) \in E(D)} \alpha(vw)}.$$

References

- [1] E. N. Barron, P. Cardaliaguet and R. Jensen, Conditional essential suprema with applications, *Appl. Math. Optim.*, **48** (2003) 229–253.
- [2] R. G. Bartle, *A Modern Theory of Integration*, American Mathematical Society, Providence, 2001.
- [3] A. Barvinok, On the number of matrices and a random matrix with prescribed row and column sums and 0-1 entries, *Adv. Math.*, **224** (2010) 316–339.
- [4] A. Barvinok and J. A. Hartigan, An asymptotic formula for the number of non-negative integer matrices with prescribed row and column sums, *Trans. Amer. Math. Soc.*, **364** (2012), 4323–4368
- [5] A. Barvinok and J. A. Hartigan, The number of graphs and a random graph with a given degree sequence, *Random Structures Alg.*, **42** (2013) 301–348.
- [6] E. R. Canfield, Z. Gao, C. S. Greenhill, B. D. McKay and R. W. Robinson, Asymptotic enumeration of correlation-immune boolean functions, *Cryptography and Communications*, **2** (2010) 111–126.
- [7] E. R. Canfield, C. Greenhill and B. D. McKay, Asymptotic enumeration of dense 0-1 matrices with specified line sums, *J. Combin. Th. Ser. A*, **115** (2008) 32–66.
- [8] E. R. Canfield and B. D. McKay, Asymptotic enumeration of dense 0-1 matrices with equal row sums and equal column sums, *Electron. J. Combin.*, (2005) **12**, #R29.
- [9] E. R. Canfield and B. D. McKay, Asymptotic enumeration of integer matrices with large equal row and column sums, *Combinatorica*, **30** (2010) 655–680.
- [10] O. Catoni, Laplace transform estimates and deviation inequalities, *Ann. I. H. Poincaré*, **39** (2003) 1–26.
- [11] O. Catoni, Laplace transform estimates and deviation inequalities, in *Lecture Notes in Mathematics 1851* (ed. J. Picard), (2004) 199–222.
- [12] S. Chatterjee, P. Diaconis and A. Sly, Random graphs with a given degree sequence, *Ann. Appl. Probab.*, **21** (2011) 1400–1435.
- [13] V. Csiszár, P. Hussami, J. Komlós, T. F. Móri, L. Rejtő and G. Tusnády, When the degree sequence is a sufficient statistic, *Act Math. Hungar.*, **134** (2012) 45–53.
- [14] Z. Gao, B. D. McKay and X. Wang, Asymptotic enumeration of tournaments with a given score sequence containing a specified digraph, *Random Structures Algorithms*, **16** (2000) 47–57.

- [15] C. Greenhill and B.D. McKay, Random dense bipartite graphs and directed graphs with specified degrees, *Random Struct. Alg.*, **35** (2009) 222–249.
- [16] N. J. Higham, Functions of Matrices, SIAM, Philadelphia, 2008.
- [17] W. Hoeffding, Probability Inequalities for sums of bounded random variables, *J. Amer. Stat. Ass.*, **38** (1963) 13–30.
- [18] B. Holmquist, Moments and cumulants of the multivariate normal distribution, *Stochastic Analysis and Applications*, **6** (1988) 273–278.
- [19] M. Isaev, Asymptotic behaviour of the number of Eulerian circuits, *Electron. J. Combin.*, **18** (2011), #219.
- [20] M. I. Isaev, Asymptotic enumeration of Eulerian circuits in graphs with strong mixing properties, *Izvestiya: Math.*, **77** (2013), 1105–1129.
- [21] M. Isaev, Asymptotic behaviour of the number of Eulerian orientations of graphs, *Math. Notes*, **93** (2013) 828–843.
- [22] M. I. Isaev, K. V. Isaeva, Asymptotic enumeration of Eulerian orientations for graphs with strong mixing properties (Russian), *Diskretn. Anal. Issled. Oper.*, **20** (2013) 40–58.
- [23] M. Isaev and B.D. McKay, On a bound of Hoeffding in the complex case, *Electron. Comm. Probab.*, **21** (2016) #14, 1–7.
- [24] M. Hasheminezhad and B.D. McKay, Combinatorial estimates by the switching method, *Contemporary Mathematics*, **531** (2010) 209–221.
- [25] L. Isserlis, On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables, *Biometrika*, **12** (1918) 134–139.
- [26] O. Kallenberg, Foundations of Modern Probability, Second Edition, Springer, 2001.
- [27] G. Kuperberg, S. Lovett and R. Peled, Probabilistic existence of regular combinatorial structures, [arXiv:1302.4295v2](https://arxiv.org/abs/1302.4295v2) (2013).
- [28] B. Laurent and P. Massart, Adaptive estimation of a quadratic functional by model selection, *Ann. Stat.*, **28** (2000) 1302–1338.
- [29] S. R. Lay, Convex Sets and their Applications, Courier Corp., 2007.
- [30] C. McDiarmid, Concentration, in Probabilistic Methods for Algorithmic Discrete Mathematics, *Algorithms and Combinatorics*, **16** (1998) 195–248.
- [31] B.D. McKay, Subgraphs of dense random graphs with specified degrees, *Combin. Probab. Comput.*, **20** (2011) 413–433.

- [32] B. D. McKay, The asymptotic numbers of regular tournaments, eulerian digraphs and eulerian oriented graphs, *Combinatorica*, **10** (1990) 367–377.
- [33] B. D. McKay and J. C. McLeod, Asymptotic enumeration of symmetric integer matrices with uniform row sums. *J. Australian Math. Soc.*, **92** (2012) 367–384.
- [34] B. D. McKay and R. W. Robinson, Asymptotic enumeration of Eulerian circuits in the complete graph, *Combin. Prob. Comput.*, **7** (1998) 437–449.
- [35] B. D. McKay and X. Wang, Asymptotic enumeration of tournaments with a given score sequence, *J. Combin. Theory Ser. A*, **73** (1996) 77–90.
- [36] B. D. McKay and N. C. Wormald, Asymptotic enumeration by degree sequence of graphs of high degree, *European J. Combin.*, **11** (1990) 565–580.
- [37] A. Montgomery, An asymptotic formula for the number of balanced incomplete block design incidence matrices, [arXiv:1407.4552](https://arxiv.org/abs/1407.4552) (2014).
- [38] E. Ordentlich and R. M. Roth, Two-dimensional weight-constrained codes through enumeration bounds, *IEEE Trans. Inform. Theory*, **46** (2000) 1292–1301.
- [39] A. Rinaldo, S. Petrović and S. E. Fienberg, Maximum likelihood estimation in the β -model, *Ann. Stat.*, **41** (2013) 1085–1110.
- [40] A. Rinaldi, S. Petrović and S. E. Fienberg, Maximum likelihood estimation in the β -model, Supplementary Materials, *Ann. Stat.*, **41** (2013) 1085–1110.
- [41] X. Wang, Asymptotic enumeration of Eulerian digraphs with multiple edges, *Australas. J. Combin.*, **5** (1992), 293–298.
- [42] X. Wang, Asymptotic enumeration of digraphs by excess sequence. Graph theory, combinatorics, and algorithms, Vol. 1, 2 (Kalamazoo, MI, 1992), 1211–1222, Wiley-Intersci. Publ., Wiley, New York, 1995.
- [43] X. Wang, The asymptotic number of Eulerian oriented graphs with multiple edges. *J. Combin. Math. Combin. Comput.*, **24** (1997) 243–248.